LECTURE 12: THE DERIVATIVE IN $\mathbb{R}^n$

1. The Derivative in $\mathbb{R}^n$

Definition: Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ and $x \in \mathbb{R}^n$

If there is a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(x+h) = f(x) + Ah + r(h) \qquad \text{with } \lim_{h \to 0} \frac{|r(h)|}{|h|} = 0$$

Then we say f is **differentiable at** x and $f'(x) = A$

Definition: f is **differentiable** if f is differentiable at all x

Note: More commonly, people write $o(h)$ instead of $r(h)$, it's a sub-linear term, smaller than h

Before, $f'(x)$ was just a number, but now it's something more dynamic, it's a linear transformation. Intuitively, if f distorts space, then $f'(x)$ describes the linear part of the distortion.

Here are a couple of immediate properties

Fact: If $f(x) = Ax$ where A is a linear transformation, then $f'(x) = A$

$$\text{Because } f(x+h) = A(x+h) = Ax + Ah = f(x) + Ah + r(h)$$

Date: Thursday, July 21, 2022.

Where $r(h) \equiv 0$, which is sublinear

Fact: If f is differentiable at x then f is continuous at x

$$\text{Follows because } f(x+h) - f(x) = f'(x)h + r(h)$$

$$\text{So } \lim_{h \to 0} f(x+h) - f(x) = \lim_{h \to 0} f'(x)h + r(h) = 0$$

Hence $\lim_{h \to 0} f(x+h) = f(x)$ $\square$

2. Uniqueness

Slightly trickier to prove is uniqueness of $f'(x)$

Fact: The derivative is unique

Proof:

STEP 1: Suppose f has two derivatives A and B at x

Then for all $y \in \mathbb{R}^n$, we have

$$\begin{aligned} f(x+y) =& f(x) + Ay + r(y) \\ f(x+y) =& f(x) + By + s(y) \end{aligned}$$

Subtracting the second equation from the first, we get

$$Ay - By = s(y) - r(y)$$

In particular, this implies that

$$\lim_{y \to 0} \frac{|Ay - By|}{|y|} = \lim_{y \to 0} \frac{|s(y) - r(y)|}{|y|} \leq \lim_{y \to 0} \frac{|s(y)| + |r(y)|}{|y|} = 0$$

STEP 2: Notice that for any $t \in \mathbb{R}$, $\lim_{t\to 0} ty = 0$ and so by the above

$$0 = \lim_{t\to 0} \frac{|A(ty) - B(ty)|}{|ty|} \overset{\text{LIN}}{=} \lim_{t\to 0} \frac{|t|\,|Ay - By|}{|t|\,|y|} = \lim_{t\to 0} \underbrace{\frac{|Ay - By|}{|y|}}_{\text{Constant}} = \frac{|Ay - By|}{|y|}$$

So in fact for every y we have

$$\frac{|Ay - By|}{|y|} = 0 \Rightarrow |Ay - By| = 0 \Rightarrow Ay = By \Rightarrow A = B \quad \square$$

3. The Chain Rule

Video: Multivariable Chain Rule

Theorem: [Chain Rule]

If f and g are differentiable and $F(x) = g(f(x))$ then

$$F'(x) = g'(f(x))f'(x)$$

Note: The right-hand-side is the composition (or matrix multiplication) of the two derivatives $g'(f(x))$ and $f'(x)$. So the derivative of the composition $g(f(x))$ is the composition of derivatives. This is what makes this formula extremely elegant.

Proof:

STEP 1: Fix x and let $A = f'(x)$ and $B = g'(f(x))$. Using the definition of F then the definition of f' and of g' we get

$$\begin{aligned} F(x+h) =& g(f(x+h)) = g(f(x) + Ah + r_f(h)) \\ =& g(f(x)) + B\left(Ah + r_f(h)\right) + r_g\left(Ah + r_f(h)\right) \end{aligned}$$

$$F(x+h) = F(x) + BAh + Br_f(h) + r_g\left(Ah + r_f(h)\right)$$

If we show that the remainder terms are sublinear, then we would be done because then

$$F'(x) = BA = g'(f(x))f'(x)$$

STEP 2: Remainder Terms

$$\frac{|Br_f(h)|}{|h|} \leq \|B\| \left(\frac{|r_f(h)|}{|h|}\right) \overset{h\to 0}{\to} 0$$

For the second term, first notice that

$$|Ah + r_f(h)| \leq \|A\| \, |h| + |r_f(h)| \overset{h\to 0}{\to} 0$$

Therefore, by definition of r_g we have

$$\frac{|r_g(Ah + r_f(h))|}{|h|} = \frac{|r_g(Ah + r_f(h))|}{|Ah + r_f(h)|} \times \frac{|Ah + r_f(h)|}{|h|} \overset{h\to 0}{\to} 0$$

This follows from the definition of r_g and because its input goes to 0, while the second term is bounded. □

Technical Note: It is possible that $Ah + r_f(h) = 0$, but this can be dealt with by redefining $r_g(0) = 0$ if necessary.

Aside: There is also an analog of the product rule $(fg)' = f'g + fg'$, but it is a bit more difficult to state, because it involves bilinear maps[1]

4. Partial Derivatives

[1]See Pugh, Chapter 5 Theorem 9(d) for details

So far the derivative $f'(x)$ we've been talking about is called the **total derivative** or the **Fréchet Derivative**.

This is in contrast with the **partial derivatives** you learned in Multivariable Calculus:

Let $\{e_1, \dots, e_n\}$ be the standard basis of $\mathbb{R}^n$ and suppose $f = (f_1, \dots, f_m)$

Definition: The $(i, j)^{th}$ partial derivative of f at x is

$$\frac{\partial f_i}{\partial x_j}(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t}$$

Note: The book writes $D_j f_i$ for $\frac{\partial f_i}{\partial x_j}$

While this definition is more natural, the total derivative is better:

Theorem: If the total derivative $f'(x)$ exists, then the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist, and in fact

$$[f'(x)] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Proof: Fix j, then since f is differentiable at x, we have

$$f(x + te_j) - f(x) = f'(x)\,(te_j) + r(te_j)$$

Where $\lim_{t \to 0} \frac{|r(te_j)|}{te_j} = \lim_{t \to 0} \frac{|r(te_j)|}{|t|} = 0$

Dividing both sides by t and using linearity of $f'(x)$ we get

$$\lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = f'(x)(e_j)$$

The left hand side is by definition $\begin{bmatrix} \frac{\partial f_1}{\partial x_j} \\ \vdots \\ \frac{\partial f_m}{\partial x_j} \end{bmatrix}$ and right-hand-side is by definition the $j-$th column of $[f'(x)]$, so both columns are equal. Since this is true for all j, the two matrices are equal. $\square$

In general, the mere existence of the partial derivatives does *not* imply that f is differentiable. But it *is* true if you assume that the partial derivatives are continuous:

Theorem: If the partial derivatives of f exist and are continuous, then f is differentiable

Proof:[2]

STEP 1:

Fix x and let A be the linear transformation whose matrix is $[A] = [\frac{\partial f_i}{\partial x_j}(x)]$. We just need to show that

$$r(h) = f(x+h) - f(x) - Ah \text{ is sublinear}$$

Because then $f'(x)$ exists and equals A

Note: From now on, assume $m = 1$, so f is a scalar function (else do the proof below with f_i instead of f)

STEP 2: Consider the path $\sigma_1 \to \sigma_2 \to \cdots \to \sigma_n$ that goes from x to $x + h$ in n straight segments (see picture in lecture)

[2]This proof is taken from Theorem 8 in Chapter 5 of Pugh's textbook

So σ_1 connects $x_0 = x$ with $x_1 = x + h_1 e_1$ and σ_2 connects x_1 with $x_2 = x_1 + h_2 e_2$, and σ_n connects x_{n-1} with $x_n = x + h$

More precisely, for each j define $\sigma_j : [0,1] \to \mathbb{R}$ as

$$\sigma_j(t) = x_{j-1} + t h_j e_j$$

$$\text{Define } g_j(t) = f(\sigma_j(t))$$

So g_j collects the value of f on the path. Notice $g_j : [0,1] \to \mathbb{R}$

Then by the Mean Value Theorem applied to g on $[0,1]$ there is $t_j \in (0,1)$ such that

$$g_j(1) - g_j(0) = g_j'(t_j)$$

But $g_j(1) = f(\sigma_j(1)) = f(x_j)$ and $g_j(0) = f(x_{j-1})$ and

$$\begin{aligned}
g_j'(t_j) &= \lim_{t \to 0} \frac{g_j(t_j + t) - g_j(t_j)}{t} \\
&= \lim_{t \to 0} \frac{f(\sigma_j(t_j + t)) - f(\sigma_j(t_j))}{t} \\
&= \lim_{t \to 0} \frac{f(\sigma_j(t_j) + t h_j e_j) - f(\sigma_j(t_j))}{t h_j} h_j \qquad \text{(using the def of } \sigma_j) \\
&= \frac{\partial f}{\partial x_j}(\sigma_j(t_j)) h_j \\
&= \frac{\partial f}{\partial x_j}(c_j) h_j \qquad \text{where } c_j =: \sigma_j(t_j)
\end{aligned}$$

Therefore the above becomes

$$f(x_j) - f(x_{j-1}) = \frac{\partial f}{\partial x_j}(c_j) h_j$$

STEP 3: Therefore we get

$$
\begin{aligned}
r(h) =& f(x+h) - f(x) - Ah \\
=& \left(\sum_{j=1}^{n} f(x_j) - f(x_{j-1}) \right) - Ah \quad \text{Telescoping Sum} \\
=& \left(\sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(c_j) h_j \right) - \left(\sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) h_j \right) \quad \text{Definition of } A \\
=& \sum_{j=1}^{n} \left[\frac{\partial f}{\partial x_j}(c_j) - \frac{\partial f}{\partial x_j}(x) \right] h_j
\end{aligned}
$$

$$
|r(h)| \leq \sum_{j=1}^{n} \left| \frac{\partial f}{\partial x_j}(c_j) - \frac{\partial f}{\partial x_j}(x) \right| |h_j| \overset{\text{C-S}}{\leq} \left(\sum_{j=1}^{n} \left| \frac{\partial f}{\partial x_j}(c_j) - \frac{\partial f}{\partial x_j}(x) \right|^2 \right)^{\frac{1}{2}} |h|
$$

$$
\frac{|r(h)|}{|h|} \leq \left(\sum_{j=1}^{n} \left| \frac{\partial f}{\partial x_j}(c_j) - \frac{\partial f}{\partial x_j}(x) \right|^2 \right)^{\frac{1}{2}}
$$

But since $c_j \to x$ as $h \to 0$ and the $\frac{\partial f}{\partial x_j}$ are continuous, this implies that the term in brackets go to 0 as $h \to 0$. Therefore we have $\lim_{h \to 0} \frac{|r(h)|}{|h|} = 0$. Hence f is differentiable at x $\square$