# Chapter 2

# Differentiation

## 2.1 Notes

1/10:

- Since manifolds look like Euclidean spaces locally, we basically only need to study differentiation on Euclidean spaces.
- Set up: Let $U \subset \mathbb{R}^n$ be open, and $f : U \to \mathbb{R}^n$ be a function.
- Idea: The derivative of $f$ at some point $\mathbf{a} \in U$ is "the best linear approximation" to $f$ at $\mathbf{a}$.
- **Differentiable** (function $f$ at $\mathbf{a}$): A function $f$ for which there exists a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ such that
$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$
- **Total derivative** (of $f$ at $\mathbf{a}$): The linear transformation $A$ corresponding to a differentiable function $f$. *Denoted by* $\boldsymbol{Df}(\mathbf{a})$.
- Questions to ask:
  1. When does the total derivative exist?
  2. When it does exist, can there be multiple?
  3. When it exists and is unique, how do I calculate it?
- Proposition: If $A, B$ are linear transformations that both satisfy the definition, then $A = B$.
  - We have
$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0} \qquad \lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - B\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$
  - It follows by subtracting the right equation above from the left one that
$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{A\mathbf{h} - B\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$
  - Apply linearity: For an arbitrary $\mathbf{v} \in \mathbb{R}^n$ and $t \in \mathbb{R}$, $t > 0$, we have
$$\frac{A(t\mathbf{v}) - B(t\mathbf{v})}{t} = A\mathbf{v} - B\mathbf{v}$$

- Therefore, since $t\mathbf{v} \to 0$ as $t \to 0$, we have by the above that

$$\begin{aligned}
\mathbf{0} &= \lim_{t\to 0} \frac{A(t\mathbf{v}) - B(t\mathbf{v})}{\|t\mathbf{v}\|} \\
&= \lim_{t\to 0} \frac{A\mathbf{v} - B\mathbf{v}}{\|\mathbf{v}\|} \\
\mathbf{0}\cdot\|\mathbf{v}\| &= \lim_{t\to 0}(A\mathbf{v} - B\mathbf{v}) \\
\mathbf{0} &= A\mathbf{v} - B\mathbf{v} \\
B\mathbf{v} &= A\mathbf{v}
\end{aligned}$$

- Example: Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be linear, i.e., $f(\mathbf{v}) = A\mathbf{v}$ for some linear transformation $A$. Then for all $\mathbf{a} \in \mathbb{R}^n$, $Df(\mathbf{a}) = A$ is constant.
  - We have from the definition that

$$\begin{aligned}
\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} &= \lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}) + f(\mathbf{h}) - f(\mathbf{a}) - f(\mathbf{h})}{\|\mathbf{h}\|} \\
&= \lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{0}}{\|\mathbf{h}\|} \\
&= \mathbf{0}
\end{aligned}$$

- Theorem: If $f$ is differentiable at $\mathbf{a}$, then $f$ is continuous at $\mathbf{a}$.
  - By definition, there exists a linear transformation $A$ such that

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

  - Additionally, we have that

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + A\mathbf{h} + \|\mathbf{h}\|\left(\frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|}\right)$$

  - As $\mathbf{h} \to \mathbf{0}$, the right-hand side of the above equation goes to $f(\mathbf{a})$.
    - As a linear transformation, $A\mathbf{h} \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$.
    - Clearly $\|\mathbf{h}\| \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$.
    - And we have by definition that the last term goes to $\mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$.
  - Therefore, $f$ is continuous at $\mathbf{a}$.
- Observation: A function $f : U \to \mathbb{R}^m$ is given by an $m$-tuple of functions $f_1 : U \to \mathbb{R}$ known as components. $f = (f_1, \dots, f_m)$.
- Proposition: $f$ is differentiable at $\mathbf{a} \in U$ iff each component function $f_i$ is differentiable at $\mathbf{a}$. In this case,

$$Df(\mathbf{a}) = (Df_1(\mathbf{a}), \dots, Df_m(\mathbf{a}))$$

  - We know that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} \in \mathbb{R}^m$$

  - Thus, the limit is zero iff the limit of each component is zero.
  - We have that the $i^{\text{th}}$ component of the vector on the left below is equal to the number on the right; we call the common value $L_i(\mathbf{h})$.

$$\left(\frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|}\right)_i = \frac{f_i(\mathbf{a}+\mathbf{h}) - f_i(\mathbf{a}) - (A\mathbf{h})_i}{\|\mathbf{h}\|}$$

- The upshot is that $f$ is differentiable at $\mathbf{a}$ iff $\lim_{\mathbf{h}\to\mathbf{0}} L_i(\mathbf{h}) = \mathbf{0}$ iff the linear transformation $\mathbf{h} \mapsto (A\mathbf{h})_i : \mathbb{R}^m \to \mathbb{R}$ is the total derivative of $f_i$.

- Now, each $f_i$ is a function of $n$ variables, i.e., $f_i(x_1, \dots, x_n)$ where $x_1, \dots, x_n$ are coordinates on $\mathbb{R}^n$.

1/12: - **Partial derivative** (of $f$ wrt. $x_i$ at $\mathbf{a} \in U$): The following quantity. *Denoted by* $\boldsymbol{\partial f/\partial x_i}$. *Given by*

$$\frac{\partial f}{\partial x_i} = \lim_{h\to 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots, a_n) - f(\mathbf{a})}{h}$$

  - The partial derivative is easy to calculate if you're good at calculating single-variable derivatives.

- Questions:
  1. If the partial derivatives all exist, does the total derivative also exist?
  2. If partial derivatives exist, is $f$ continuous?

- The answer is no to both — it's too weak a condition.
  - Counter example: Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^4} & (x, y) \neq \mathbf{0} \\ 0 & (x, y) = \mathbf{0} \end{cases}$$

  - All partial derivatives exist at $(0, 0)$ but $f$ is not continuous at $(0, 0)$.
  - We'll consider this in the homework.

- Now we try taking derivatives in infinitely many directions, as opposed to just $n$ many.

- **Directional derivative** (of $f$ at $\mathbf{a}$ in the direction of $\mathbf{v} \in \mathbb{R}^n$): The following quantity. *Denoted by* $\boldsymbol{D_{\mathbf{v}} f(\mathbf{a})}$, $\boldsymbol{\partial f/\partial \mathbf{v}}$. *Given by*

$$D_{\mathbf{v}} f(\mathbf{a}) = \frac{\partial f}{\partial \mathbf{v}} = \lim_{h\to 0} \frac{f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a})}{h}$$

  - We always take $\|\mathbf{v}\| = 1$.
  - The partial derivative is just a directional derivative along the standard basis vectors. Alternatively, the directional derivative is just a generalization of the partial derivatives.

- This still isn't a strong enough condition — the above counterexample has all directional derivatives at $(0, 0)$ but still isn't continuous.

- Proposition: Suppose $f$ is differentiable at $\mathbf{a} \in U$. Then all directional derivatives of $f$ at $\mathbf{a}$ exist and for all $\mathbf{v} \in \mathbb{R}^n$,

$$\frac{\partial f}{\partial \mathbf{v}} = Df(\mathbf{a})(\mathbf{v})$$

  - The total derivative says that the derivative exists from all sequences of approach. We're just going to pick a particular vector direction of approach.
  - Mathematically, by the definition of the total derivative,

$$\begin{aligned} \mathbf{0} &= \lim_{h\to 0} \frac{f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a}) - Df(\mathbf{a})(h\mathbf{v})}{h} \\ &= \lim_{h\to 0} \frac{f(\mathbf{a} + h\mathbf{v}) - f(\mathbf{a})}{h} - Df(\mathbf{a})(\mathbf{v}) \\ Df(\mathbf{a})(\mathbf{v}) &= \frac{\partial f}{\partial \mathbf{v}} \end{aligned}$$

- A particular consequence is that
$$\frac{\partial f}{\partial x_i} = Df(\mathbf{a})(e_i)$$
  - But the total derivative, as a linear transformation, is completely defined by its behavior on the basis vectors.
  - Thus, it is defined by the $m$-by-$n$ matrix
$$Df(\mathbf{a}) = \left(\frac{\partial f_j}{\partial x_i}\right)_{\substack{1\le j\le m\\ 1\le i\le n}}$$
- **Jacobian matrix** (of $f$ at $\mathbf{a}$): The above matrix, representing the total derivative of $f$ at $\mathbf{a}$.
- Theorem: Suppose $f : U \to \mathbb{R}^m$ is a function on an open set $U \subset \mathbb{R}^n$. If all partial derivatives of $f$ exist and are continuous on $U$, then $f$ is differentiable on $U$.
  - Recall the mean value theorem (MVT): Suppose $g : [a, b] \to \mathbb{R}$ is a continuous function which is differentiable on $(a, b)$. Then there exists $c \in (a, b)$ such that $g'(c) = [g(b) - g(a)]/[b - a]$.
  - WLOG let $m = 1$ (if we prove this case, we can use the proposition relating $f$ to its components to prove the general case).
  - Rewrite
$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) &= f(a_1+h_1, a_2+h_2, a_3+h_3, \dots, a_n+h_n) - f(a_1, a_2+h_2, a_3+h_3, \dots, a_n+h_n) \\ &\quad + f(a_1, a_2+h_2, a_3+h_3, \dots, a_n+h_n) - f(a_1, a_2, a_3+h_3, \dots, a_n+h_n) \\ &\quad + \cdots \\ &\quad + f(a_1, \dots, a_{n-1}, a_n+h_n) - f(\mathbf{a}) \end{aligned}$$
    where $\mathbf{a} = (a_1, \dots, a_n)$ and $\mathbf{h} = (h_1, \dots, h_n)$.
  - Apply the MVT to each term to get
$$f(a_1, \dots, a_i+h_i, \dots, a_n+h_n) - f(a_1, \dots, a_i, \dots, a_n+h_n) = h_i \frac{\partial f}{\partial x_i}(a_1, \dots, c_i(\mathbf{h}), \dots, a_n+h_n)$$
    for some $c_i(\mathbf{h}) \in (a_i, a_i+h_i) \cup (a_i+h_i, a_i)$.
  - Now let $A$ be the Jacobian matrix of $f$ at $\mathbf{a}$.
  - WTS:
$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} = \mathbf{0}$$
  - We have
$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(a_1, \dots, c_i(\mathbf{h}), \dots, a_n+h_n)$$
  - Let $\pi_i : \mathbb{R}^n \to \mathbb{R}^n$ be the linear map $(x_1, \dots, x_n) \mapsto (0, \dots, x_i, \dots, 0)$. Clearly, $\mathbf{x} = \sum_{i=1}^n \pi_i \mathbf{x}$.
  - Thus, $A\mathbf{h} = \sum_{i=1}^n A\pi_i\mathbf{h}$ and $A\pi_i\mathbf{h} = \frac{\partial f}{\partial x_i}(\mathbf{a}) \cdot h_i$.
  - Applying, we have
$$\begin{aligned} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - A\mathbf{h}}{\|\mathbf{h}\|} &= \sum_{i=1}^{n} \frac{1}{\|\mathbf{h}\|}\left(h_i \frac{\partial f}{\partial x_i}(a_1, \dots, a_{i-1}, c_i(\mathbf{h}), a_{i+1}+h_{i+1}, \dots, a_n+h_n) - \frac{\partial f}{\partial x_i}(\mathbf{a}) \cdot h_i\right) \\ &= \sum_{i=1}^{n} \frac{h_i}{\|\mathbf{h}\|}\left(\frac{\partial f}{\partial x_i}(a_1, \dots, a_{i-1}, c_i(\mathbf{h}), a_{i+1}+h_{i+1}, \dots, a_n+h_n) - \frac{\partial f}{\partial x_i}(\mathbf{a})\right) \end{aligned}$$
  - We know that $-1 \le h_i/\|\mathbf{h}\| \le 1$, so we need only show that the difference above goes to zero as $\mathbf{h} \to \mathbf{0}$. But we know this by the continuity of the partial derivatives.

- Note that this theorem gives a sufficient condition but not a necessary condition for $f$ to be differentiable.

1/14: - Theorem (Chain Rule): Suppose $f : U \to \mathbb{R}^m$ and $g : V \to \mathbb{R}^p$ are functions defined on open sets $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ with $f(U) \subset V$. Suppose that $f$ is differentiable at $\mathbf{a} \in U$ and $g$ is differentiable at $\mathbf{b} = f(\mathbf{a}) \in V$. Then the composite function $g \circ f : U \to \mathbb{R}^p$ is differentiable at $\mathbf{a}$ and $D(g \circ f)(\mathbf{a}) = Dg(\mathbf{b}) \circ Df(\mathbf{a}) : \mathbb{R}^n \to \mathbb{R}^p$.
  - Note that $f : U \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in U$ with derivative $A$ iff $f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + A\mathbf{h} + \tilde{f}(\mathbf{h})$ such that
$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\tilde{f}(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0}$$
    where $\tilde{f}$ is an error function.
    - We're just rearranging terms here.
    - If you like, $\tilde{f}$ is the numerator from the definition of the total derivative.
  - Let $A = Df(\mathbf{a})$, $B = Dg(\mathbf{b})$. Then
$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + A\mathbf{h} + \tilde{f}(\mathbf{h})$$
    so
$$\begin{aligned}(g \circ f)(\mathbf{a}+\mathbf{h}) &= g(f(\mathbf{a}+\mathbf{h}))\\ &= g(f(\mathbf{a})) + A\mathbf{h} + \tilde{f}(\mathbf{h})\\ &= g(f(\mathbf{a})) + B(A\mathbf{h} + \tilde{f}(\mathbf{h})) + \tilde{g}(A\mathbf{h} + \tilde{f}(\mathbf{h}))\\ &= g(f(\mathbf{a})) + BA\mathbf{h} + B\tilde{f}(\mathbf{h}) + \tilde{g}(A\mathbf{h} + \tilde{f}(\mathbf{h}))\end{aligned}$$
  - WTS: $\lim_{\mathbf{h}\to\mathbf{0}}[B\tilde{f}(\mathbf{h}) + \tilde{g}(A\mathbf{h} + \tilde{f}(\mathbf{h}))]/\|\mathbf{h}\| = \mathbf{0}$.
  - For the first half of the fraction,
$$\frac{B\tilde{f}(\mathbf{h})}{\|\mathbf{h}\|} = B\left(\frac{\tilde{f}(\mathbf{h})}{\|\mathbf{h}\|}\right) \to \mathbf{0}$$
    as $\mathbf{h} \to \mathbf{0}$ since the argument goes to $\mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$ and $B$ is a linear transformation (in particular, $B(\mathbf{0}) = \mathbf{0}$).
  - For the second half of the fraction,
$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\tilde{g}(A\mathbf{h} + \tilde{f}(\mathbf{h}))}{\|\mathbf{h}\|} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{\tilde{g}(A\mathbf{h} + \tilde{f}(\mathbf{h}))}{\left\|A\mathbf{h} + \tilde{f}(\mathbf{h})\right\|} \cdot \frac{\left\|A\mathbf{h} + \tilde{f}(\mathbf{h})\right\|}{\|\mathbf{h}\|}$$
    - The left fraction on the right side of the equality goes to zero as $\mathbf{h} \to \mathbf{0}$ by the definition of $\tilde{g}$.
    - The right fraction on the right side of the equality is bounded since
$$\frac{\left\|A\mathbf{h} + \tilde{f}(\mathbf{h})\right\|}{\|\mathbf{h}\|} \leq \frac{\|A\mathbf{h}\|}{\|\mathbf{h}\|} + \frac{\left\|\tilde{f}(\mathbf{h})\right\|}{\|\mathbf{h}\|} \leq \|A\| + \frac{\left\|\tilde{f}(\mathbf{h})\right\|}{\|\mathbf{h}\|}$$
      where $\|A\|$ is the operator norm and $\left\|\tilde{f}(\mathbf{h})\right\|/\|\mathbf{h}\| \to 0$ as $\mathbf{h} \to \mathbf{0}$ by the definition of $\tilde{f}$.
    - Thus, the second half of the fraction goes to zero as well.
- Theorem: Let $U \subset \mathbb{R}^m$ be an open subset.

1. Suppose $f, g : U \to \mathbb{R}^m$ are functions that are differentiable at $\mathbf{a} \in U$. Then $f + g$ is also differentiable at $\mathbf{a} \in U$ and
$$D(f+g)(\mathbf{a}) = Df(\mathbf{a}) + Dg(\mathbf{a})$$
2. Suppose $f, g : U \to \mathbb{R}$ are both differentiable at $\mathbf{a} \in U$. Then $f \cdot g$ is also differentiable at $\mathbf{a}$, and
$$D(f \cdot g)(\mathbf{a}) = Df(\mathbf{a}) \cdot g(\mathbf{a}) + f(\mathbf{a}) \cdot Dg(\mathbf{a})$$
3. Suppose $f : U \to \mathbb{R}$ is differentiable at $\mathbf{a} \in U$ and $f(\mathbf{a}) \neq 0$. Then $1/f$ is differentiable at $\mathbf{a} \in U$ and
$$D(1/f)(\mathbf{a}) = -\frac{Df(\mathbf{a})}{f(\mathbf{a})^2}$$

– Proof of 1: Consider the functions $F : U \to \mathbb{R}^{2m}$ and $G : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ defined by
$$F(\mathbf{x}) = (f(\mathbf{x}), g(\mathbf{x})) \qquad G(\mathbf{y}, \mathbf{z}) = \mathbf{y} + \mathbf{z}$$
so that
$$f + g = G \circ F$$
  - $F$ is differentiable because its components are differentiable.
  - $G$ is differentiable because it's linear. This also implies that $DG(\mathbf{x}) = G$.
  - Apply the chain rule to learn that $G \circ F$ is differentiable with derivative
$$\begin{aligned} D(f+g)(\mathbf{a}) &= D(G \circ F)(\mathbf{a}) \\ &= DG(F(\mathbf{a})) \circ DF(\mathbf{a}) \\ &= G(DF(\mathbf{a})) \\ &= G(Df(\mathbf{a}), Dg(\mathbf{a})) \\ &= Df(\mathbf{a}) + Dg(\mathbf{a}) \end{aligned}$$

– Prove the others the same way.

- Theorem (Mean Value Theorem): Suppose $f : U \to \mathbb{R}$ is differentiable for all $\mathbf{a} \in U$ and that $U$ contains the line segment joining $\mathbf{a}, \mathbf{a} + \mathbf{h} \in U$. Then there exists $t_0 \in (0, 1)$ such that
$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = Df(\mathbf{a} + t_0\mathbf{h})(\mathbf{h})$$
  – Define $\phi(t) = f(\mathbf{a} + t\mathbf{h})$ for $t \in [0, 1]$.
  – Apply the usual MVT to $\phi$ to learn that there exists $t_0 \in (0, 1)$ such that $\phi(1) - \phi(0) = \phi'(t_0)$.
  – Then using the chain rule, $\phi'(t_0) = Df(\mathbf{a} + t_0\mathbf{h})(\mathbf{h})$.
- We now discuss higher order derivatives.
- **Differentiable** ($f$ on $U$): A function $f$ that is differentiable at every $\mathbf{a} \in U$.
- If $f$ is differentiable on $U$, then the total derivative gives a map $Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.
  – Note that $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ is isomorphic to the set of all $m$-by-$n$ matrices, and $\mathbb{R}^{mn}$.
- We can ask for $Df$ to itself be differentiable. We define
$$D^2 f = D(Df)$$
if it exists and, more generally,
$$D^k f = D(D^{k-1} f)$$
- **Class $C^k$** (function): A function $f : U \to \mathbb{R}^m$ for which $Df, \ldots, D^k f$ all exist and are continuous on $U$.

  - Note that we technically need only require that $D^k f$ exist, as this implies the existence of $Df, \dots, D^{k-1}f$.
  - A function $f : U \to \mathbb{R}^m$ is of class $C^k$ iff all partial derivatives $\partial f/\partial x_i : U \to \mathbb{R}^m$ exist and are of class $C^{k-1}$ (this follows from the theorem relating partial derivatives and differentiability).

- **Smooth** (function): A function of class $C^\infty$.

1/19:

- Theorem: Let $U \subset \mathbb{R}^n$ be open, and let $f : U \to \mathbb{R}^m$ be a $C^2$ function. Then for any $i, j$,

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

  - WLOG, let $n = 2$ (because only two variables play a role in this theorem; we don't need the others) and $m = 1$ (we can do this for each component separately).
  - See the figure/proof associated with Theorem 15.3 in Labalme (2021).

- Note that there is a homework problem giving an example of a function for which the above quantities are defined but not equal. This doesn't violate the theorem, though, because the function isn't $C^2$.
- Next goal: One of the most important theorems in this class — the inverse function theorem.
- Theorem (Inverse function theorem): Suppose $U \subset \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^n$ is a $C^k$ function for $k \geq 1$, and $\mathbf{a} \in U$ such that $Df(\mathbf{a}) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible. Then there exist open neighborhoods $V \subset U$ of $\mathbf{a}$ and $W$ of $\mathbf{b} = f(\mathbf{a}) \in \mathbb{R}^n$ and a $C^k$ function $g : W \to V$ such that $f(V) = W$, $g(W) = V$, $g(f(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in V$, and $f(g(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in W$. Moreover, $Dg(\mathbf{y}) = Df(g(\mathbf{y}))^{-1}$ for all $y \in W$.
  - The proof will be reasonably involved because we have to deal with the construction of the open neighborhoods. The proof will be reasonably nonconstructive?
  - Intuition: If $n = 1$, $f'(\mathbf{a}) \neq 0$ implies that $f$ is increasing or decreasing in a neighborhood of $\mathbf{a}$; in either case, there's a local inverse.
    - In higher dimensions, the derivative won't be up or down but will look like a linear transformation.
  - Idea of proof: Given $\mathbf{y} \in \mathbb{R}^n$ near $\mathbf{b} = f(\mathbf{a})$, we want to find $\mathbf{x} \in U$ near $\mathbf{a}$ such that $f(\mathbf{x}) = \mathbf{y}$.
    - First guess: Take $\mathbf{x}_0 = \mathbf{a} - A^{-1}(\mathbf{b} - \mathbf{y})$ where $A = Df(\mathbf{a})$.
    - Hope: $f(\mathbf{x}_0)$ is closer to $\mathbf{y}$ than $\mathbf{b} = f(\mathbf{a})$.
    - Then we iterate; do this again and again to get a sequence that converges to the point we want. In particular, $\mathbf{x}_1 = \mathbf{x}_0 - A^{-1}(f(\mathbf{x}_0) - \mathbf{y})$, and on and on.
    - We're going to formalize this idea of iteration using the **contraction mapping theorem** (aka the Banach fixed point theorem).
    - Let $\mathbf{y} \in \mathbb{R}^n$ be fixed near $\mathbf{b} = f(\mathbf{a})$. $F_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} - A^{-1}(f(\mathbf{x}) - \mathbf{y})$. Note: $f(\mathbf{x}) = \mathbf{y}$ iff $F_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$.
    - Goal: Find a fixed point of $F_{\mathbf{y}}(\mathbf{x})$ (another way of iterating the sequence is that at some point you will get a fixed point, i.e., a point that you can plug into the recursion relation and get the same point back out).

- Theorem (Contraction mapping theorem): Let $X$ be a complete metric space, and suppose $T : X \to X$ is a function for which there exists a constant $r < 1$ such that for all $\mathbf{x}, \mathbf{y} \in X$, $d(T(\mathbf{x}), T(\mathbf{y})) \leq r \cdot d(\mathbf{x}, \mathbf{y})$. Then $T$ has a unique fixed point.
- The way you find a fixed point is by starting with an arbitrary point and then just iterating $T$ to it. Then show that in the limit it converges to something, and that something is a fixed point.
- You can't have more than one fixed point.

## 2.2 Chapter 2: Differentiation

*From Munkres (1991).*

1/18:

- **Directional derivative** (of $f$ at $\mathbf{a}$ with respect to $\mathbf{u}$): The following limit, where $A \subset \mathbb{R}^m$ contains a neighborhood of $\mathbf{a}$, $f : A \to \mathbb{R}^n$, and $\mathbf{u} \in \mathbb{R}^m$ is nonzero. *Denoted by* $\boldsymbol{f'(\mathbf{a};\mathbf{u})}$. *Given by*

$$f'(\mathbf{a};\mathbf{u}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{u}) - f(\mathbf{a})}{t}$$

  - Note that it is not necessary for $\mathbf{u}$ to be a unit vector.
- If we choose as our definition of differentiability "$f$ is differentiable at $\mathbf{a}$ if $f'(\mathbf{a};\mathbf{u})$ exists for every $\mathbf{u} \neq \mathbf{0}$," we would not have results such as differentiability implies continuity and the chain rule.
  - Thus, we need a stronger definition.
- As an alternate definition of differentiability in the one-variable case, consider the following.
- **Differentiable** (single-variable real function at $a$): A function $\phi : A \to \mathbb{R}$, where $A \subset \mathbb{R}$ contains a neighborhood of $a$, for which there exists a number $\lambda$ such that

$$\frac{\phi(a+t) - \phi(a) - \lambda t}{t} \to 0 \quad \text{as} \quad t \to 0$$

- **Derivative** (of a single-variable real function at $a$): The unique number $\lambda$ in the above definition. *Denoted by* $\phi'(a)$.
- "This formulation of the definition makes explicit the fact that if $\phi$ is differentiable, then the linear function $\lambda t$ is a good approximation to the **increment function** $\phi(a+t) - \phi(a)$; we often call $\lambda t$ the **first-order approximation** or **linear approximation** to the increment function" (Munkres, 1991, p. 43).
- **Increment function**: The function $\phi(a+t) - \phi(a)$.
- **First-order approximation**: The function $\lambda t$. *Also known as* **linear approximation**.
- To generalize the idea of a first-order/linear approximation to the increment function $f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a})$, we take a function that is linear in the sense of linear algebra.
- Note that either the sup norm or the Euclidean norm can be used in the definition of the total derivative.
- Theorem 5.1: Let $A \subset \mathbb{R}^m$, and let $f : A \to \mathbb{R}^n$. If $f$ is differentiable at $\mathbf{a}$, then all the directional derivatives of $f$ at $\mathbf{a}$ exist, and

$$f'(\mathbf{a};\mathbf{u}) = Df(\mathbf{a}) \cdot \mathbf{u}$$

- Theorem 5.2: Let $A \subset \mathbb{R}^m$, and let $f : A \to \mathbb{R}^n$. If $f$ is differentiable at $\mathbf{a}$, then $f$ is continuous at $\mathbf{a}$.
- $\boldsymbol{j}^{\textbf{th}}$ **partial derivative** (of $f$ at $\mathbf{a}$): The directional derivative of $f$ at $\mathbf{a}$ with respect to the vector $\mathbf{e}_j$, provided this derivative exists. *Denoted by* $\boldsymbol{D_j f(\mathbf{a})}$.
- Theorem 5.3. Let $A \subset \mathbb{R}^m$, and let $f : A \to \mathbb{R}$. If $f$ is differentiable at $\mathbf{a}$, then

$$Df(\mathbf{a}) = \begin{bmatrix} D_1 f(\mathbf{a}) & \cdots & D_m f(\mathbf{a}) \end{bmatrix}$$

- Theorem 5.4: Let $A \subset \mathbb{R}^m$, and let $f : A \to \mathbb{R}^n$. Suppose $A$ contains a neighborhood of $\mathbf{a}$. Let $f_i : A \to \mathbb{R}$ be the $i^{\text{th}}$ component function of $f$ so that

$$f(\mathbf{x}) = \begin{bmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{bmatrix}$$

(a) The function $f$ is differentiable at $\mathbf{a}$ if and only if each component function $f_i$ is differentiable at $\mathbf{a}$.

(b) If $f$ is differentiable at $\mathbf{a}$, then its derivative is the $n$-by-$m$ matrix whose $i^{\text{th}}$ row is the derivative of the function $f_i$, i.e.,

$$Df(\mathbf{a}) = \begin{bmatrix} Df_1(\mathbf{a}) \\ \vdots \\ Df_n(\mathbf{a}) \end{bmatrix}$$

or, in other words, $Df(\mathbf{a})$ is the matrix whose entry in row $i$ and column $j$ is $D_j f_i(\mathbf{a})$.

- "It is possible for the partial derivatives, and hence the Jacobian matrix, to exist *without* it following that $f$ is differentiable at $\mathbf{a}$" (Munkres, 1991, p. 47).
  - As per the example outlined in class.
- Special cases ($m = 1$ or $n = 1$).
  - If $f : \mathbb{R}^1 \to \mathbb{R}^3$, $f$ is often interpreted as a parameterized curve and

    $$Df(t) = \begin{bmatrix} f_1'(t) \\ f_2'(t) \\ f_3'(t) \end{bmatrix}$$

    is the velocity vector of the curve.
  - If $g : \mathbb{R}^3 \to \mathbb{R}^1$, $g$ is often interpreted as a scalar field, and the vector field

    $$Dg(\mathbf{x}) = \begin{bmatrix} D_1 g(\mathbf{x}) & D_2 g(\mathbf{x}) & D_3 g(\mathbf{x}) \end{bmatrix}$$

    is called the gradient of $g$.
    - In this case, the directional derivative of $g$ with respect to $\mathbf{u}$ is written in calculus as the dot product of the vectors $\vec{\nabla} g$ and $\mathbf{u}$.
- Theorem 6.1 (Mean Value Theorem): If $\phi : [a, b] \to \mathbb{R}$ is continuous at each point of the closed interval $[a, b]$, and differentiable at each point of the open interval $(a, b)$, then there exists a point $c$ of $(a, b)$ such that

  $$\phi(b) - \phi(a) = \phi'(c)(b - a)$$

- Theorem 6.2: Let $A$ be open in $\mathbb{R}^m$. Suppose that the partial derivatives $D_j f_i(\mathbf{x})$ of the component functions of $f$ exist at each point $\mathbf{x}$ of $A$ and are continuous on $A$. Then $f$ is differentiable at each point of $A$.
- **Continuously differentiable** (function on $A$): A function $f$ for which the partial derivatives $D_j f_i(\mathbf{x})$ of the component functions of $f$ exist at each point $\mathbf{x} \in A$ and are continuous on $A$, where $A \subset \mathbb{R}^m$ is open. *Also known as* **class $\boldsymbol{C^1}$** (function on $A$).
- There are differentiable functions that are not of class $C^1$, but we will not concern ourselves with them.
- Theorem 6.3[1]: Let $A$ be open in $\mathbb{R}^m$, and let $f : A \to \mathbb{R}$ be a function of class $C^2$ on $A$. Then for each $\mathbf{a} \in A$,

  $$D_k D_j f(\mathbf{a}) = D_j D_k f(\mathbf{a})$$

- Theorem 7.1 (Chain Rule): Let $A \subset \mathbb{R}^m$, $B \subset \mathbb{R}^n$, $f : A \to \mathbb{R}^n$, $g : B \to \mathbb{R}^p$, $f(A) \subset B$, and $\mathbf{b} = f(\mathbf{a})$. If $f$ is differentiable at $\mathbf{a}$ and $g$ is differentiable at $\mathbf{b}$, then the composite function $g \circ f$ is differentiable at $\mathbf{a}$. Furthermore,

  $$D(g \circ f)(\mathbf{a}) = Dg(\mathbf{b}) \cdot Df(\mathbf{a})$$

  where the indicated product is matrix multiplication.

[1] See Theorem 15.3 in Labalme (2021).

- Corollary 7.2: Let $A$ be open in $\mathbb{R}^m$, and let $B$ be open in $\mathbb{R}^n$. Let $f : A \to \mathbb{R}^n$ and $g : B \to \mathbb{R}^p$ with $f(A) \subset B$. If $f$ and $g$ are of class $C^r$, so is the composite function $g \circ f$.
- Theorem 7.3 (Mean Value Theorem): Let $A$ be open in $\mathbb{R}^m$, and let $f : A \to \mathbb{R}$ be differentiable on $A$. If $A$ contains the line segment with end points $\mathbf{a}$ and $\mathbf{a} + \mathbf{h}$, then there is a point $\mathbf{c} = \mathbf{a} + t_0\mathbf{h}$ with $0 < t_0 < 1$ of this line segment such that

$$f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) = Df(\mathbf{c}) \cdot \mathbf{h}$$

- Theorem 7.4: Let $A$ be open in $\mathbb{R}^n$, $f : A \to \mathbb{R}^n$, and $\mathbf{b} = f(\mathbf{a})$. Suppose that $g$ maps a neighborhood of $\mathbf{b}$ into $\mathbb{R}^n$, that $g(\mathbf{b}) = \mathbf{a}$, and $g(f(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x}$ in a neighborhood of $\mathbf{a}$. If $f$ is differentiable at $\mathbf{a}$ and if $g$ is differentiable at $\mathbf{b}$, then

$$Dg(\mathbf{b}) = [Df(\mathbf{a})]^{-1}$$

*Proof.* Let $i : \mathbb{R}^n \to \mathbb{R}^n$ be the identity function. It has total derivative $I_n$. But since $g(f(\mathbf{x})) = i(\mathbf{x})$ for all $\mathbf{x}$ in a neighborhood of $\mathbf{a}$, the Chain Rule implies that

$$\begin{aligned}
Dg(\mathbf{b}) \cdot Df(\mathbf{a}) &= I_n \\
Dg(\mathbf{b}) &= [Df(\mathbf{a})]^{-1}
\end{aligned}$$

as desired. □

- It follows from Theorem 7.4 that for $f^{-1}$ to be differentiable at $\mathbf{a}$, it is *necessary* that $Df(\mathbf{a})$ is invertible.
  - We will later prove that this condition is also *sufficient* for a function $f$ of class $C^1$ to have a differentiable inverse.
- **Functional notation**: Notation such as $\phi'$ for a derivative.
- **Operator notation**: Notation such as $D\phi$ for a derivative.
- Munkres (1991) argues that Leibniz notation is a relic of a "time when the focus of every physical and mathematical problem was on the *variables* involved, and when *functions* as such were hardly even thought about" (p. 60).