Calculus

Khanh Nguyen

May 2023

As a physics student in high school, CS undergraduate in NTU, I never have a proper/formal introduction to calculus. This notes is my best effort to formalize calculus for real numbers and served as a cheat sheet as well

1 Derivative

Let $f \in \mathcal{C}^1(\mathbb{R}^m, \mathbb{R}^n)$. In this notes, to simplify the analysis, we assume f being continuously differentiable.

We define the derivative $D : \mathcal{C}^1(\mathbb{R}^m, \mathbb{R}^n) \to \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ [1] where

- $\mathcal{C}^1(\mathbb{R}^m, \mathbb{R}^n)$: set of all continuously differentiable functions from $\mathbb{R}^m$ to $\mathbb{R}^n$
- $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$: set of all linear maps from $\mathbb{R}^m$ to $\mathbb{R}^n$

The derivative D is an operator acting on the set of all continuously differentiable functions from $\mathbb{R}^m$ to $\mathbb{R}^n$ and produce a linear map on every points on $\mathbb{R}^m$

The derivative of f at $x_0 \in \mathbb{R}^m$ is the *closest* linear map [2] to f at x_0

$$f(x) = f(x_0) + Df(x_0)(x - x_0) + o(x - x_0) \tag{1}$$

where $o(x - x_0) \subset \mathbb{R}^m \to \mathbb{R}^n$ is the set of functions strictly smaller than $||x - x_0||$ when x approaches x_0, i.e $\lim_{x \to x_0} \frac{||o(x-x_0)||}{||x-x_0||} = 0$

$Df(x)$ is a linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ which is isomorphic to the Jacobian matrix $J_{f(x)} \in \mathbb{R}^{n \times m}$

$$J_{f(x)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & ... & \frac{\partial f_1}{\partial x_m} \\ ... & ... & ... \\ \frac{\partial f_n}{\partial x_1} & ... & \frac{\partial f_n}{\partial x_m} \end{bmatrix} \tag{2}$$

[1] **function currying notation**
$A \to B$: set of all functions from A to B
$A \to B \to C = A \to (B \to C)$

[2] the existence and uniqueness of $Df(x_0)$ is a corrolary of Taylor's Theorem

2 Linearity of Derivative Operator

The derivative operator is linear. Given $f, g \in \mathcal{C}^1(\mathbb{R}^m, \mathbb{R}^n)$ and $\alpha \in \mathbb{R}$

- $D(f + g) = Df + Dg$
- $D\alpha f = \alpha Df$

where the addition and scalar multiplication are defined as $(f + g)(x) = f(x) + g(x)$ and $(\alpha f)(x) = \alpha f(x)$

3 Derivative of element-wise function

Given a scalar function $\sigma \in \mathcal{C}^1(\mathbb{R}, \mathbb{R})$ with derivate $D\sigma(x) \in \mathcal{L}(\mathbb{R}, \mathbb{R})$ which is a 1d linear function, i.e $y = ax$ for $a, x \in \mathbb{R}$

We can generalize this function to n-dimensional $\bar{\sigma}(x) \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^n)$ such that if $y = \bar{\sigma}(x)$ for $x \in \mathbb{R}^n$, then we have

$$y_i = \sigma(x_i) \tag{3}$$

We can calculate the Jacobian matrix of this function

$$J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & ... & \frac{\partial y_1}{\partial x_n} \\ ... & ... & ... \\ \frac{\partial y_n}{\partial x_1} & ... & \frac{\partial y_n}{\partial x_n} \end{bmatrix} \tag{4}$$

where

$$J_{i,j} = \begin{cases} D\sigma(x_i) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{5}$$

Jacobian matrix is diagonal

$$J = \begin{bmatrix} D\sigma(x_0) & ... & 0 \\ ... & ... & ... \\ 0 & ... & D\sigma(x_n) \end{bmatrix} \tag{6}$$

4 Chain Rule

Chain rule is used when we can to calculate the derivative of a function composition. Let's derive the chain rule.

Given $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^k)$, $g \in \mathcal{C}^1(\mathbb{R}^m, \mathbb{R}^n)$, and $f \circ g \in \mathcal{C}^1(\mathbb{R}^m, \mathbb{R}^k)$. We will calculate $D(f \circ g)$ given Df and Dg

Let $x, x_0 \in \mathbb{R}^m$ and $y = g(x), y_0 = g(x_0) \in \mathbb{R}^n$

$Df(y_0)$ is the derivative of f at y_0, we have

$$f(y) = f(y_0) + Df(y_0)(y - y_0) + o(y - y_0) \tag{7}$$

Rewrite

$$(f \circ g)(x) = (f \circ g)(x_0) + Df(g(x_0))(g(x) - g(x_0)) + o(g(x) - g(x_0)) \tag{8}$$

$Dg(x_0)$ is the derivative of f at x_0, we have

$$g(x) = g(x_0) + Dg(x_0)(x - x_0) + o(x - x_0) \tag{9}$$

Hence

$$\begin{aligned}(f \circ g)(x) &= (f \circ g)(x_0) + Df(g(x_0))(Dg(x_0)(x - x_0) + o(x - x_0)) + o(g(x) - g(x_0)) \\ &= (f \circ g)(x_0) + Df(g(x_0))Dg(x_0)(x - x_0) + Df(g(x_0))o(x - x_0) + o(g(x) - g(x_0))\end{aligned}$$

We can prove $Df(g(x_0))o(x-x_0)+o(g(x)-g(x_0)) = o(x-x_0)$ by assuming $||Df(g(x_0))|| < \infty$ and $||Dg(x_0)|| < \infty$

$Df(g(x_0))o(x - x_0)$ is small

$$||Df(g(x_0))o(x - x_0)|| \le ||Df(g(x_0))||||o(x - x_0)|| \qquad \text{(operator norm)}$$

We have

$$\begin{aligned}\lim_{x \to x_0} \frac{||Df(g(x_0))o(x - x_0)||}{||x - x_0||} &\le ||Df(g(x_0))|| \lim_{x \to x_0} \frac{||o(x - x_0)||}{||x - x_0||} \\ &= 0\end{aligned}$$

$o(g(x) - g(x_0))$ is small

$$\begin{aligned}\frac{||o(g(x) - g(x_0))||}{||x - x_0||} &= \frac{||o(g(x) - g(x_0))||}{||g(x) - g(x_0)||} \frac{||g(x) - g(x_0)||}{||x - x_0||} \\ &= \frac{||o(g(x) - g(x_0))||}{||g(x) - g(x_0)||} \frac{||Dg(x_0)(x - x_0) + o(x - x_0)||}{||x - x_0||} \\ &\le \frac{||o(g(x) - g(x_0))||}{||g(x) - g(x_0)||} \frac{||Dg(x_0)(x - x_0)|| + ||o(x - x_0)||}{||x - x_0||} && \text{(Cauchy-Schwarz)} \\ &\le \frac{||o(g(x) - g(x_0))||}{||g(x) - g(x_0)||} \frac{||Dg(x_0)||||(x - x_0)|| + ||o(x - x_0)||}{||x - x_0||} && \text{(operator norm)} \\ &= \frac{||o(g(x) - g(x_0))||}{||g(x) - g(x_0)||} (||Dg(x_0)|| + \frac{||o(x - x_0)||}{||x - x_0||})\end{aligned}$$

We have

$$\lim_{x\to x_0} \frac{||o(g(x)-g(x_0))||}{||x-x_0||} \leq \lim_{x\to x_0} \frac{||o(g(x)-g(x_0))||}{||g(x)-g(x_0)||}(||Dg(x_0)|| + \frac{||o(x-x_0)||}{||x-x_0||}$$
$$= 0$$

Therefore,

$$(f\circ g)(x) = (f\circ g)(x_0) + Df(g(x_0))Dg(x_0)(x-x_0) + o(x-x_0) \tag{10}$$

Theorem 1 (Chain Rule)**.** *Given two differentiable function $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^k)$, $g \in \mathcal{C}^1(\mathbb{R}^m, \mathbb{R}^n)$, the derivative of the composition $f \circ g$ is $D(f\circ g)(x) = Df(g(x))Dg(x)$*

5 Change of variables

Given a Lebesgue integral on a measurable subset $E \subset \mathbb{R}^n$ of positive measurable function $f : E \to [0, \infty)$

$$\int_E f \tag{11}$$

Let $G : \Omega \to E$ be a parameterization satisfying certain conditions [3] in Ω. Then

$$\int_E f = \int_\Omega (f\circ G)|DG| \tag{12}$$

where $|DG|(x) = |(DG)(x)|$ is the determinant of derivative of G

6 Higher Order Derivative

In first-order derivative, Df is defined as a set of linear maps on every point in the domain $\mathbb{R}^m$

$$f \in \mathbb{R}^m \to \mathbb{R}^n$$
$$Df \in \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$$

Notice that $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ is isomorphic to $\mathbb{R}^{n\times m}$ which is also a finite dimensional vector space on $\mathbb{R}$. And all finite dimensional vector space on $\mathbb{R}$ is isomorphic to $\mathbb{R}^d$ for some $d \in \mathbb{N}$. We can define the second-order derivative of f as follows

[3] $\mathcal{C}^1$ diffeomorphism

$$D^2 f = DDf \in \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)) = \mathbb{R}^m \to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^m; \mathbb{R}^n) \tag{13}$$

where $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^m; \mathbb{R}^n)$ denotes the set of all bi-linear map from $\mathbb{R}^m \times \mathbb{R}^m$ to $\mathbb{R}^n$ [4]

7 Algebra on Space of Operator

Let $A, B, C \in (X \to Y) \to (X \to Y)$, linear operator $D \in (X \to Y) \to (X \to Y)$, $f \in X \to Y$, define

- Addition: $(A + B)f = Af + Bf$
- Commutativity of addition: $A + B = B + A$
- Associativity of addition: $(A + B) + C = A + (B + C)$
- Composition: $(AB)f = A(Bf)$
- Associativity of composition: $(AB)C = A(BC)$
- Distributivity: $(A + B)C = AC + BC$
- Distributivity for linear operator: $D(A + B) = DA + DB$

7.1 Eigenfunction

Let Y be a vector space over a field $\mathbb{K}$

An eigenfunction of a operator $A \in (X \to Y) \to (X \to Y)$ is $f \in X \to Y$ such that

$$Af = \lambda f \tag{14}$$

for some scalar $\lambda \in \mathbb{K}$ and define operator λ as $(\lambda f)(x) = \lambda f(x)$

7.2 Polynomial of linear operator

Let Y be a vector space over a field $\mathbb{K}$, $r_1, ..., r_n \in \mathbb{K}$

We can expand the products of linear factors of linear operator D as follows

$$(D + r_1)(D + r_2)...(D + r_n) = a_0 + a_1 D + a_2 D^2 + ... + a_n D^n \tag{15}$$

The RHS is a polynomial of degree n of D

Proof by induction: Let $\lambda \in \mathbb{K}$, we have the commutativity of scalar operator $\lambda D = D\lambda$

[4] proof will be done until my mind is clearer

$$\begin{aligned}
(D+\lambda)\sum_{j=0}^{n} a_j D^j &= D\sum_{j=0}^{n} a_j D^j + \lambda \sum_{j=0}^{n} a_j D^j && \text{(distributivity)}\\
&= \sum_{j=0}^{n} D a_j D^j + \lambda \sum_{j=0}^{n} a_j D^j && \text{(distributivity of linear operator)}\\
&= \sum_{j=0}^{n} a_j D^{j+1} + \lambda \sum_{j=0}^{n} a_j D^j && \text{(commutativity of scalar operator)}\\
&= \lambda a_0 + (\sum_{j=1}^{n} (a_{j-1} + \lambda a_j) D^j) + a_n D^{n+1} && \text{(communtative, associative of addition)}
\end{aligned}$$

Example: Given $f \in \mathbb{R} \to \mathbb{R}$, linear $D \in (\mathbb{R} \to \mathbb{R}) \to (\mathbb{R} \to \mathbb{R})$, $(1f)(x) = f(x)$, $(2f)(x) = 2f(x)$

$$\begin{aligned}
(D+1)(D+2) &= D(D+2) + 1(D+2) && \text{(distributivity)}\\
&= D^2 + D2 + 1D + 1 \cdot 2 && \text{(distributivity for linear operator)}\\
&= D^2 + D2 + 1D + 2 && \text{(property of scalar operator)}\\
&= D^2 + 2D + 1D + 2 && \text{(property of scalar operator)}\\
&= D^2 + (2+1)D + 2 && \text{(distributivity)}\\
&= D^2 + 3D + 2 && \text{(property of scalar operator)}
\end{aligned}$$

7.3 Differential equation with constant coefficients

Let $a_0, a_1, ..., a_n \in \mathbb{R}$ and $x, y \in \mathcal{C}^n(\mathbb{R}, \mathbb{R})$

$$a_0 y + a_1 y' + a_2 y'' + a_3 y^{(3)} + ... + a_n y^{(n)} = x \tag{16}$$

Or in the differential operator form

$$(a_0 + a_1 D + a_2 D^2 + ... + a_n D^n) y = x \tag{17}$$

Suppose we can write the polynomial of D as product of linear factors [5]

$$(D - r_1)(D - r_2)...(D - r_n) y = x \tag{18}$$

Let $z \in \mathbb{R} \to \mathbb{R}$ be the function produced by the last $n-1$ linear factors applied on y, i.e $z = (D - r_2)...(D - r_n) y$ [6]. We have

[5] in Real field, every polynomial can be written as product of linear and quadratic factors, in Complex field, every polynomial can be written as product of linear factors. Here I tried to make it as simple as possible. The idea is more important

[6] z exists because y is n times continuously differentiable

$$(D - r_1)z = x \tag{19}$$

This is a constant-coefficient first-order non-homogeneous linear ODE.

8 Hessian Matrix

The second-order derivative of $f \in \mathbb{R}^n \to \mathbb{R}$ is

$$D^2 f \in \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^n, \mathbb{R})) \tag{20}$$

A linear map from $\mathbb{R}^n$ to its dual space is isomorphic to an $n \times n$ matrix, hence Hessian matrix. The second derivative can be represented as

$$D^2 f(x) = x^T H_{f(x)} x \tag{21}$$

where $H_{f(x)}$ is the Hessian matrix of f at x

$$H_{f(x)} = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial f}{\partial^2 x_n} \end{bmatrix} \tag{22}$$

Another note is that H is symmetric by the symmetric property of second-derivative, so its eigenvalues are real. Another nice property, H is positive semi-definite everywhere on $\mathbb{R}^n$ if and only if f is convex.

9 Green's Theorem - Stoke's Theorem - Divergence Theorem - Generalized Stoke's Theorem

10 Differential Forms