VV285 RC Part V

Differential Calculus

First Derivative, Regulated Integral

Xingjian Zhang

Univerity of Michigan-Shanghai Jiao Tong University Joint Institute

June 20, 2020

Outline

- Big- and Small-"O" Notation
- Derivative of a Function
- Partial Derivative
- Jacobian
- * Second Derivative
- Product Rule
- Chain Rule
- Integral of Step Functions
- Mean Value Theorem

Something you need to pay attention to...

Think More and Be Interactive!

- Do think more about the question in “()”.
  e.g. “(How to prove?)”
- You are welcome to ask questions in a adequate manner.
- Please open your camera so that I can receive more feedbacks from you. (Makes our life easier!)
- The class is designed to be interactive. However, if you really do not want to be asked at all, please type an “_” before your zoom name.

Big- and Small-"O" Notation

Landau Symbols:

Let $f : X \to V_1, g : X \to V_2$ and $x_0 \in X$. We say that

$$f(x) = o(g(x)) \qquad \text{as } x \to x_0 \qquad \Leftrightarrow \qquad \lim_{x \to x_0} \frac{\|f(x)\|_{V_1}}{\|g(x)\|_{V_2}} = 0$$

and

$$f(x) = O(g(x))$$

has analogous definition as VV186.

Remark The meaning of $f(x) = o(g(x))$: $f(x)$ is significantly less than $g(x)$. The meaning of $f(x) = O(g(x))$: $f(x)$ is not significantly greater than $g(x)$.

Derivative of a Function

Let X, V be finite-dimensional vector spaces and $\Omega \subset X$ an open set. Then a map $f : \Omega \to V$ is called *differentiable at $x \in \Omega$* if there exists a linear map $L_x \in \mathcal{L}(X, V)$ such that

$$f(x+h) = f(x) + L_x h + o(h) \qquad \text{as } h \to 0. \tag{1}$$

In this case we call L_x the *derivative of f at x* and write

$$L_x = Df|_x = df|_x.$$

We say that f is differentiable on Ω if it is differentiable for every $x \in \Omega$.

(How to prove derivative is well-defined? i.e. prove its uniqueness)

Exercise

Let f be a function and $f : X \to V$. Distinguish these maps by clarifying it maps from which vector space to which vector space. Are they linear?

1. D,
2. Df,
3. $Df|_x$
4. D^2
5. $D^2 f$,

What about $D^2 f|_x$, $D^2 f|_x y$...

Derivative of Linear Map

The derivative of a linear map L at some point is L itself. (Why?)

Example:

1. $Df|_z(h) = \overline{h}$, where $f(z) = \overline{z}$
2. $DA|_x(h) = Ah$
3. $D\,\mathrm{tr}|_A H = \mathrm{tr}\, H$
4. $DD|_f = Df$

Remark: For $A \in \mathrm{GL}(n; \mathbb{C})$, you will prove (in Ex 5.1.(iii)) that

$$D\det|_A H = \det A \,\mathrm{tr}\left(A^{-1}H\right)$$

(Recall that determinant is multi-linear, why it's derivative is not equal to itself?)

Exercise

Let A be a matrix.

Easy:
Calculate the derivative of $f(A) = A^3$.

Hard:
Prove the derivative of matrix inverse is: $D(\cdot)^{-1}|_A H = -A^{-1}HA^{-1}$.
(Hint: Start with $(A+H)^{-1} = (A(\mathrm{id} + A^{-1}H))^{-1}$.)

The Derivative Is in the Form of Matrix

If exists,

$$Df|_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \cong \mathrm{Mat}(m \times n; \mathbb{R}).$$

How to obtain this matrix? Denote by e_j the jth standard basis vector in $\mathbb{R}^n$ or $\mathbb{R}^m$. We consider the columns of $Df|_x$, which are given by $Df|_x e_j$, $j = 1, \ldots, n$. Furthermore, the (i,j)th element of $Df|_x$ is given by $\langle\, e_i\,,\, Df|_x e_j \,\rangle$. (Recall *matrix elements* in *Exercise 2.2*!)

Partial Derivative

Let $\Omega \subset \mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$ be differentiable on Ω. We then define the *partial derivative with respect to x_j at $x \in \Omega$* by

$$\left.\frac{\partial f}{\partial x_j}\right|_x := \lim_{h\to 0} \frac{f(x+he_j) - f(x)}{h}$$

$$= \lim_{h\to 0} \frac{f(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(x)}{h}$$

In this notation,

$$(Df|_x)_{ij} = \frac{\partial f_i}{\partial x_j}$$

or rather

$$Df|_x = \left(\frac{\partial f}{\partial x_1} \cdots \frac{\partial f}{\partial x_n}\right) = \left.\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}\right|_x$$

Jacobian

Let $\Omega \subset \mathbb{R}^n$ and $f : \Omega \to \mathbb{R}^m$. Assume that all partial derivatives $\frac{\partial f_i}{\partial x_j}$ of f exist at $x \in \Omega$. The matrix

$$J_f(x) := \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \Bigg|_x$$

called the *Jacobian* of f.

If the derivative $Df|_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ exists, $J_f(x) \in \mathrm{Mat}(m \times n; \mathbb{R})$ is the representing matrix of $Df|_x$ w.r.t. the standard bases in $\mathbb{R}^n$ and $\mathbb{R}^m$.

Remark: However, the existence of Jacobian $J_f(x)$ does not imply the differentiability of f at x.

Partial Derivatives and Continuity

1. All partial derivatives are bounded $\Rightarrow$ f is *continuous*;
2. All partial derivatives are continuous $\Rightarrow$ f is *continuously differentiable*.

Remark: Recall the difference between definition of continuous, differentiable, and continuously differentiable. A function can be differentiable while not continuously differentiable!
There is a gap between these two results. We now have two simple ways to sufficiently deduce whether f is continuous or continuously differentiable at x. However, there is no simple way for us to conclude f is or not differentiable at x.

Example I

Let's look at this example, where partial derivatives are not all continuous and f is still differentiable. Define a function $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = \begin{cases} (x^2+y^2)\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right), & \text{if } (x,y) \neq (0,0) \\ 0, & \text{if } (x,y) = (0,0) \end{cases}$$

$$\left.\frac{\partial f}{\partial x}\right|_{(0,0)} = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = 0$$

$$\left.\frac{\partial f}{\partial x}\right|_{(x,y)\neq(0,0)} = 2x\sin\left(\frac{1}{\sqrt{x^2+y^2}}\right) - \frac{x}{\sqrt{x^2+y^2}}\cos\left(\frac{1}{\sqrt{x^2+y^2}}\right)$$

$\frac{\partial f}{\partial x}$ exists on $\mathbb{R}^2$, but is not continuous on (0,0).

Example II

We will verify whether f is differentiable on (0,0) or not with $Df|_{(0,0)} = (0,0)$. For $h = (h_1, h_2)^T$

$$f(h) = f(0) + Df|_{(0,0)} h + (h_1^2 + h_2^2) \sin\left(\frac{1}{\sqrt{h_1^2 + h_2^2}}\right)$$

$$\lim_{h\to 0} \frac{\left|(h_1^2 + h_2^2) \sin\left(\frac{1}{\sqrt{h_1^2 + h_2^2}}\right)\right|}{\|h\|} = \lim_{h\to 0} \|h\| \cdot \left|\sin\left(\frac{1}{\sqrt{h_1^2 + h_2^2}}\right)\right|$$

which implies that

$$(h_1^2 + h_2^2) \sin\left(\frac{1}{\sqrt{h_1^2 + h_2^2}}\right) = o(h)$$

Therefore, f is differentiable on (0,0).

* Second Derivative (for Potential Function)

Extension: The second derivative of a potential function is in the form of a matrix called *Hessian*. And

$$\operatorname{Hess} f(x) = D(\nabla f)|_x = \begin{pmatrix} \left.\frac{\partial^2 f}{\partial x_1 \partial x_1}\right|_x & \left.\frac{\partial^2 f}{\partial x_2 \partial x_1}\right|_x & \cdots & \left.\frac{\partial^2 f}{\partial x_n \partial x_1}\right|_x \\ \vdots & \vdots & & \vdots \\ \left.\frac{\partial^2 f}{\partial x_1 \partial x_n}\right|_x & \left.\frac{\partial^2 f}{\partial x_2 \partial x_n}\right|_x & \cdots & \left.\frac{\partial^2 f}{\partial x_n \partial x_n}\right|_x \end{pmatrix}$$

where $\nabla f(x) = (Df|_x)^T$. How does Hessian work? We see that if $\tilde{h} \in \mathbb{R}^n$ is some other vector, $D^2 f|_x h$ acts on $\tilde{h}$ via

$$(D^2 f|_x h)\tilde{h} = (\operatorname{Hess} f(x) h)^T \tilde{h} = \langle \operatorname{Hess} f(x) h \,,\, \tilde{h} \rangle \in \mathbb{R}.$$

Furthermore, by Schwarz's Theorem, if $f \in C^2(\Omega, V)$, $\operatorname{Hess} f(x) = \operatorname{Hess} f(x)^T$. i.e.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Product Rule

Let U, X_1, X_2, V be finite-dimensional vector spaces and $\Omega \subset U$ an open set. Let $f : \Omega \to X_1$ and $g : \Omega \to X_2$ be differentiable maps and $\odot : X_1 \times X_2 \to V$ a *generalized product*. Then $f \odot g : \Omega \to V$ is also differentiable and

$$D(f \odot g) = (Df) \odot g + f \odot (Dg). \tag{2}$$

At $x \in \Omega$ the right-hand side is interpreted as a linear map $U \to V$

$$u \mapsto D(f \odot g)|_x u = (Df|_x u) \odot g(x) + f(x) \odot (Dg|_x u). \tag{3}$$

Question: What is the derivative of $k(t) = f(t) \times g(t)$?

Chain Rule

Let U, X, V be finite-dimensional vector spaces and $\Omega \subset U, \Sigma \subset X$ open sets. Let $g : \Omega \to \Sigma$ and $f : \Sigma \to V$ be differentiable maps. Then the composition $f \circ g : \Omega \to V$ is also differentiable and for all $x \in \Omega$

$$D(f \circ g)|_x = Df|_{g(x)} \circ Dg|_x, \tag{4}$$

where the right-hand side is a composition of linear maps.

The proof is basically identical to that of 186 Theorem 3.1.12, the chain rule for functions of one real variable.

Exericise

Calculate the derivative of $\text{tr}(AA^T)$ in two ways:

1. By definition;
2. By chain rule and product rule.

Integral of Step Functions I

Let $I \subset \mathbb{R}$ be an interval and $(V, \|\cdot\|_V)$ a normed vector space. We say that a map $f : I \to V$ is *bounded* if

$$\|f\|_\infty := \sup_{x \in I} \|f(x)\|_V < \infty. \tag{5}$$

The set of all bounded functions $f : I \to V$ is denoted $L^\infty(I, V)$.

A sequence of functions (f_n), $f_n : I \to V$, $I \subset \mathbb{R}$, *converges uniformly* to $f : I \to V$ in a normed vector space $(V, \|\cdot\|_V)$ if

$$\|f_n - f\|_\infty := \sup_{x \in I} \|f_n(x) - f(x)\|_V \xrightarrow{x \to \infty} 0.$$

A function is *regulated* if it is the uniform limit of a sequence of step functions. (What is the closure of the set of step functions in the uniform norm?)

Integral of Step Functions II

The standard estimate

$$\left|\int_a^b f(x)dx\right|_V \le \int_a^b \|f(x)\|_V dx \le |b-a| \cdot \sup_{x\in[a,b]} \|f(x)\|_V.$$

is useful in some proof. You can consider it as a generalization of triangular inequality.

What are we going to integrate? Vector-Valued functions. For now, however, we are integrating $f : [a, b] \to \mathbb{R}^n$ and

$$\int_a^b f(x)dx = \int_a^b \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix} dx = \begin{pmatrix} \int_a^b f_1(x)dx \\ \vdots \\ \int_a^b f_n(x)dx \end{pmatrix}$$

Later, we will define the integral of general vector-valued function.

Mean Value Theorem I

Let X, V be finite-dimensional vector spaces, $\Omega \subset X$ open and $f \in C^1(\Omega, V)$. Let $x, y \in \Omega$ and assume that the line segment $x + ty, 0 \le t \le 1$, is wholly contained in Ω. Then

$$f(x+y) - f(x) = \int_0^1 Df|_{x+ty} y dt = \left(\int_0^1 Df|_{x+ty} dt \right) y. \qquad (6)$$

The integrals in (6) are integrals of elements of V (the integrand $Df|_{x+ty}y$) and $\mathcal{L}(X, V)$ (the integrand $Df|_{x+ty}$) (NOT trivial!). Here, the Mean Value Theorem can be understood as a generalization of the fundamental theorem of calculus. i.e.

$$f(x+y) - f(x) = \int_x^{x+y} f'(\xi) d\xi.$$

(How?)

Mean Value Theorem II

Mean Value Theorem is NOT trivial. See this example:
$f : (x_1, x_2) \mapsto x_1^2 + x_2^2$. We calculate $f(x+y) - f(x)$ in three ways given by Mean Value Theorem.

Derivative Estimate

From the standard estimate and the Mean Value Theorem, we have

$$\|f(x+y)-f(x)\|_V \le \|y\|_X \cdot \sup_{0\le t\le 1} \|Df|_{x+ty}\|,$$

where $\|Df|_{x+ty}\|$ denotes the operator norm of $Df|_{x+ty} \in \mathcal{L}(X,V)$.

Differentiating under an Integral

Let X, V be finite-dimensional vector spaces, $I = [a, b] \subset \mathbb{R}$ an interval and $\Omega \subset X$ an open set. Let $f : I \times \Omega \to V$ be a continuous function such that $Df(t, \cdot)$ exists and is continuous for every $t \in I$. Then

$$g(x) = \int_a^b f(t, x)dt$$

is differentiable in Ω and

$$Dg(x) = \int_a^b Df(t, \cdot)|_x dt$$

Recap of Euler Gamma Function I

An application of previous result is *Euler Gamma Function*. We look into

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx$$

which can be obtained by repeated integration by parts starting from the formula

$$\int_0^\infty e^{-x} dx = 1$$

when $n = 0$. Now we are going to derive Euler's formula by repeated differentiation after introducing a parameter t. For $t > 0$, let $x = tu$. Then $dx = tdu$ and the above equation becomes

$$\int_0^\infty te^{-tu} du = 1 \Leftrightarrow \int_0^\infty e^{-tu} du = \frac{1}{t} \qquad (*)$$

Recap of Euler Gamma Function II

We need $t > 0$ in order that e^{-tx} is integrable over the region $x \geq 0$. Now, let's differentiate $(*)$ iteratively, and substitute u with x:

$$\int_0^\infty -xe^{-tx}dx = -\frac{1}{t^2}$$
$$\int_0^\infty -x^2e^{-tx}dx = -\frac{2}{t^3}$$
$$\int_0^\infty x^3e^{-tx}\mathrm{d}x = \frac{6}{t^4}$$
$$\int_0^\infty x^4e^{-tx}\mathrm{d}x = \frac{24}{t^5}$$
$$\int_0^\infty x^5e^{-tx}\mathrm{d}x = \frac{120}{t^6}$$
$$\int_0^\infty x^ne^{-tx}\mathrm{d}x = \frac{n!}{t^{n+1}}$$

Recap of Euler Gamma Function III

Let $t = 1$, we find:

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx = (n-1)!$$

About Assignment 5

(5.1) Derivative of Determinant

(5.2) Application of Chain Rule

(5.3) Second Derivative of a Potential Function (You can verify Schwarz's Theorem by this question).

(5.4) Application of Product Rule

(5.7) Dirichlet Integral (In VV286, we will learn brand-new techniques, *complex analysis*, to solve the same question!)

(5.8) The Condition of Theorem 12.9.

Have Fun

And

Learn Well!