A FIRST COURSE

IN

ANALYSIS

A FIRST COURSE

IN

ANALYSIS

MAT2006 Notebook

Lecturer

Prof. Weiming Ni

The Chinese University of Hongkong, Shenzhen

Tex Written By

Mr. Jie Wang

The Chinese University of Hongkong, Shenzhen

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Contents

Acknowledgments

This book is taken notes from the MAT2006 in fall semester, 2018. These lecture notes were taken and compiled in LaTeX by Jie Wang, an undergraduate student in Fall 2018. Prof. Weiming Ni has not edited this document. Students taking this course may use the notes as part of their reading and reference materials. This version of the lecture notes were revised and extended for many times, but may still contain many mistakes and typos, including English grammatical and spelling errors, in the notes. It would be greatly appreciated if those students, who will use the notes as their reading or reference material, tell any mistakes and typos to Jie Wang for improving this notebook.

Notations and Conventions

$\mathbb{R}^n$	n-dimensional real space
$\mathbb{C}^n$	n-dimensional complex space
$\mathbb{R}^{m\times n}$	set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m\times n}$	set of all $m \times n$ complex-valued matrices
x_i	ith entry of column vector $\boldsymbol{x}$
a_{ij}	(i,j)th entry of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i$	ith column of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i^{\mathrm{T}}$	ith row of matrix $\boldsymbol{A}$
$\mathbb{S}^n$	set of all $n \times n$ real symmetric matrices, i.e., $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $a_{ij} = a_{ji}$ for all i,j
$\mathbb{H}^n$	set of all $n \times n$ complex Hermitian matrices, i.e., $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ and $\bar{a}_{ij} = a_{ji}$ for all i,j
$\boldsymbol{A}^{\mathrm{T}}$	transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}$ means $b_{ji} = a_{ij}$ for all i,j
$\boldsymbol{A}^{\mathrm{H}}$	Hermitian transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{H}}$ means $b_{ji} = \bar{a}_{ij}$ for all i,j
$\mathrm{trace}(\boldsymbol{A})$	sum of diagonal entries of square matrix $\boldsymbol{A}$
$\mathbf{1}$	A vector with all 1 entries
$\mathbf{0}$	either a vector of all zeros, or a matrix of all zeros
$\boldsymbol{e}_i$	a unit vector with the nonzero element at the ith entry
$\mathcal{C}(\boldsymbol{A})$	the column space of $\boldsymbol{A}$
$\mathcal{R}(\boldsymbol{A})$	the row space of $\boldsymbol{A}$
$\mathcal{N}(\boldsymbol{A})$	the null space of $\boldsymbol{A}$
$\mathrm{Proj}_{\mathcal{M}}(\boldsymbol{A})$	the projection of $\boldsymbol{A}$ onto the set $\mathcal{M}$

Chapter 9

Week9

9.1. Friday

We are now in the multi-variate differentiate part.

Comments on question in last lecture. The question left in the last lecture is that

> In dimension $\mathbb{R}^k$ with k to be determined, is it possible to find the smallest k such that a sphere S^2 and a circle S^1 have a way of putting to make each point from S^2 to S^1 have the same distance?

The answer is $k = 5$, Let's give a proof. We define the sphere to be

$$S^2 = \{(x, y, z, 0, 0) \mid x^2 + y^2 + z^2 = 1\},$$

and the circle is

$$S^1 = \{(0, 0, 0, u, v) \mid u^2 + v^2 = 1\}$$

The distance between any two points on the sphere and the circile, respectively, is

$$d = \sqrt{x^2 + y^2 + z^2 + u^+ v^2} = \sqrt{2}$$

Why $k \leq 4$ is not ok?

9.1.1. Preliminaries

Define $\boldsymbol{x} := (x_1, \ldots, x_m)$ and $\boldsymbol{y} := (y_1, \ldots, y_m)$, we define the L_2 norm

$$\begin{aligned} d(\boldsymbol{x}, \boldsymbol{y}) :&= \|\boldsymbol{x} - \boldsymbol{y}\| \\ &= \left[(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2\right]^{1/2} \end{aligned}$$

and L_1, L_∞ (sup) norm:

$$\begin{aligned} d_1(\boldsymbol{x}, \boldsymbol{y}) :&= \|\boldsymbol{x} - \boldsymbol{y}\| := |x_1 - y_1| + \cdots + |x_m - y_m| \\ d_S(\boldsymbol{x}, \boldsymbol{y}) :&= \|\boldsymbol{x} - \boldsymbol{y}\|_S := \max_{1 \le i \le m} (x_i - y_i) \end{aligned}$$

Definition 9.1 [Norm] A norm $\|\cdot\|$ is a function from a vector space X to $\mathbb{R}$ such that

1. $\|\boldsymbol{x}\| \ge 0$, $\forall x \in X$; and $\|\boldsymbol{x}\| = 0$ iff $\boldsymbol{x} = \boldsymbol{0}$;
2. $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$, for $\forall x \in X, \lambda \in \mathbb{R}$;
3. $\|\boldsymbol{x}_1 + \boldsymbol{x}_2\| \le \|\boldsymbol{x}_1\| + \|\boldsymbol{x}_2\|$

R Any norm defines a metric: $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$. We (by defalut) pre-assume the norm to be L_2 norm and the metric to be L_2 metric.

However, the norm does not define the angle between two angles. So we introduce the concept for inner product:

Definition 9.2 [Inner Product] An **inner product** is a bi-operation function $\langle \cdot, \cdot \rangle : X \times X \mapsto \mathbb{R}$ such that

1. $\langle x, x \rangle \ge 0$ for $\forall x \in X$ and $\langle x, x \rangle = 0$ iff $\boldsymbol{x} = 0$
2. $\in \boldsymbol{x}\boldsymbol{y} = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$, for $\forall \boldsymbol{x}, \boldsymbol{y} \in X$
3. $\langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for $\forall x, y \in X$ and $\forall \lambda \in \mathbb{R}$
4. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

R

1. Any inner product always defines a norm, i.e., $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$.
2. The angle θ between vector $\boldsymbol{x}, \boldsymbol{y}$ is defined as:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos\theta$$

   In particular, $\theta = 0$ implies $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

9.1.2. Differentiation

Review for One-dimension. Given a function $f : \mathbb{R} \mapsto \mathbb{R}$, the derivative is defined as

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

or equivalently,

$$\begin{aligned}
0 &= \lim_{x \to x_0} \left[\frac{f(x) - f(x_0)}{x - x_0} - f'(x_0) \right] \\
&= \lim_{x \to x_0} \frac{f(x) - f(x_0) - f'(x_0)(x - x_0)}{x - x_0} \\
&= \lim_{x \to x_0} \left| \frac{f(x) - [f(x_0) + f'(x_0)(x - x_0)]}{x - x_0} \right|
\end{aligned}$$

The interpretation is that this affine (linear function) $f(x_0) + f'(x_0)(x - x_0)$ approximate f near x_0 in at least first order. This is similar to the way we define the high-dimension derivative.

High-Dimension Derivative.

Definition 9.3 [Differentiable] A map $f : U \mapsto \mathbb{R}^m$, where U is open in $\mathbb{R}^m$, is **differentiable** at $\boldsymbol{x}_0 \in U$ if

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \frac{\|f(\boldsymbol{x}) - f(\boldsymbol{x}_0) - \boldsymbol{L}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)\|}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} = 0 \tag{9.1}$$

R Note that $f(\boldsymbol{x}), f(\boldsymbol{x}_0) \in \mathbb{R}^n$, but $(\boldsymbol{x} - \boldsymbol{x}_0) \in \mathbb{R}^m$; thus $L(\boldsymbol{x}_0)$ is a **linear transformation** from $\mathbb{R}^m$ to $\mathbb{R}^n$, i.e., a $n \times m$ matrix. In this course, $\boldsymbol{L}(\boldsymbol{x}_0)$ is denoted as $Df(\boldsymbol{x}_0)$, or sometimes denoted as $f'(\boldsymbol{x}_0)$.

Interpretation. We re-write $f(\boldsymbol{x})$ as:

$$f(\boldsymbol{x}) = \begin{pmatrix} f_1(\boldsymbol{x}) & \cdots & f_n(\boldsymbol{x}) \end{pmatrix}$$
$$f_i : \mathbb{R}^m \mapsto \mathbb{R}$$

Let's study only one component first, i.e., $n = 1$. To make life simpler, let's set $m = 2$:

cut curve $y - x_1$; $y - x_2$. study the multi-variate differentiate. For fixed x_2.

$$\frac{\|f(x_1 + h, x_2) - f(x_1, x_2) - \text{Number} * h\|}{\|(h, 0)\|}$$

such a number is defined as $\frac{\partial f}{\partial x_1}(\boldsymbol{x}_0)$, with $\boldsymbol{x}_0 = (x_1, x_2)$.

Similarly, we define

$$\frac{\partial f}{\partial x_2}(\boldsymbol{x}_0),$$

i.e., the partial derivatives of f at (x_1, x_2).

Corollary 9.1 f is differentiable at $\boldsymbol{x}_0$ implies all partial derivatives of f at $\boldsymbol{x}_0$ exist.

The converse is not true. Let's raise a counter-example to explain that.

■ **Example 9.1**

$$f(x_1, x_2) = \begin{cases} \dfrac{x_1 x_2}{x_1^2 + x_2^2}, & (x_1, x_2) \neq (0, 0) \\ 0, & (x_1, x_2) = (0, 0) \end{cases}$$

When $x_2 = m x_1$, we have

$$f(x_1, m x_1) = \frac{m x_1^2}{x_1^2 + m x_1^2} = \frac{m}{1 + m^2},$$

i.e., f is not differentiable at the origin.

However, $\frac{\partial f}{\partial x_1}(0,0) = 0 = \frac{\partial f}{\partial x_2}(0,0)$.

Geometric meaning for $n = 1$: tangent plane.

What guarntees f to be differntiable if all partial derivatives exist?

$$\frac{\|f(\boldsymbol{x}) - f(\boldsymbol{x}_0) - Df(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0)\|}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} \to 0,$$

where $f : \mathbb{R}^m \mapsto \mathbb{R}^n$ and $Df(\boldsymbol{x}_0)$ be a $n \times m$ matrix. Note that we wrtie $f(\boldsymbol{x}_0)$ in row vector form (which is convenient as will be seen in future):

$$\begin{pmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_n(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} f_1(\boldsymbol{x}_0) \\ f_2(\boldsymbol{x}_0) \\ \vdots \\ f_n(\boldsymbol{x}_0) \end{pmatrix} + [Df(\boldsymbol{x}_0)]_{n\times m}(\boldsymbol{x} - \boldsymbol{x}_0)_{m\times 1} + o(\|\boldsymbol{x} - \boldsymbol{x}_0\|),$$

where $o(\|\boldsymbol{x} - \boldsymbol{x}_0\|)$ is a $m \times 1$ vector that has order less than $\|\boldsymbol{x} - \boldsymbol{x}_0\|$.

Or we write in this form:

$$f_1(\boldsymbol{x}) = f_1(\boldsymbol{x}_0) + \nabla f_1(\boldsymbol{x}_0) \cdot (\boldsymbol{x} - \boldsymbol{x}_0) + o(\|\boldsymbol{x} - \boldsymbol{x}_0\|),$$

with $\nabla f_1(\boldsymbol{x}_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}_0) & \cdots & \frac{\partial f_1}{\partial x_m}(\boldsymbol{x}_0) \end{pmatrix}$ to be a row vector, and $(\boldsymbol{x} - \boldsymbol{x}_0)$ to be a column vector.

Inverse function theorem; implicit function theorem.

Sufficient Condition for differentiability. Recall that continus have nowhere differentble point. The gap is continuous differentiable.

Theorem 9.1 Let $f : U \mapsto \mathbb{R}^n$, where $U \subseteq \mathbb{R}^m$ is open. If all partial derivatives of f are **continuous** in U, then f is differentiable in U.

Proof.

$$
\begin{aligned}
f(\boldsymbol{x}+\boldsymbol{h}) - f(\boldsymbol{x}) &= f(x_1+h_1,\ldots,x_m+h_m) - f(x_1,\ldots,x_m) \\
&= [f(x_1+h_1,\ldots,x_m+h_m) - f(x_1,x_2+h_2,\ldots,x_m+h_m)] \\
&\quad\quad + [f(x_1,x_2+h_2,\ldots,x_m+h_m) - f(x_1,x_2,x_3+h_3,\ldots,x_m+h_m)] \\
&\quad\quad + \cdots + [f(x_1,x_2,\ldots,x_m+h_m) - f(x_1,x_2,\ldots,x_m)] \\
&= \frac{\partial f}{\partial x_1}(x_1,x_2+h_2,\ldots,x_m+h_m) + o(h_1) \\
&\quad + \frac{\partial f}{\partial x_2}(x_1,x_2,\ldots,x_m+h_m)h_2 + o(h_2) \\
&\quad + \cdots + \\
&= \frac{\partial f}{\partial x_1}(x_1,\ldots,x_m)h_1 + o(1)h_1 + o(h_1) \\
&\quad + \frac{\partial f}{\partial x_2}(x_1,\ldots,x_m)h_2 + o(1)h_2 + o(h_2) + \cdots
\end{aligned}
$$

■

The next lecture will talk about the chain rule, the inverse, the derivative of inverse, the directional derivative, the gradient, the lattice curve. Next Friday will talk about implicit function theorem.