X2TE2109: Convex Optimization and Its Applications in Signal Processing

Handout B: Real Analysis Cheat Sheet

Instructor: Anthony Man–Cho So Updated: January 11, 2022

The purpose of this handout is to give a brief review of some of the basic concepts and results in analysis. If you are not familiar with the material and/or would like to do some further reading, you may consult, e.g., the books [1, 4, 3, 2].

1 Basic Topology

In this course, we shall frequently operate on various metric spaces. Recall that a metric space is a pair (X, d), where X is a set and $d : X \times X \to \mathbb{R}$ is a function satisfying the following properties:

(a) **(Non–Negativity)** for any $x, y \in X$, $d(x, y) > 0$ if $x \neq y$ and $d(x, x) = 0$;

(b) **(Symmetry)** for any $x, y \in X$, $d(x, y) = d(y, x)$;

(c) **(Triangle Inequality)** for any $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

One of the most important examples of a metric space is $(\mathbb{R}^n, d)$, where $n \geq 1$ is an integer and $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is given by

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{1/2}.$$

To facilitate our subsequent discussion, let us begin with some definitions. Let (X, d) be a metric space. The **open ball** with center at $\bar{x} \in X$ and radius $r > 0$ is defined as the set

$$B^\circ(\bar{x}, r) = \{ x \in X : d(x, \bar{x}) < r \}.$$

In a similar fashion, the **closed ball** with center at $\bar{x} \in X$ and radius $r > 0$ is defined as the set

$$B(\bar{x}, r) = \{ x \in X : d(x, \bar{x}) \leq r \}.$$

Now, for any point $x \in X$ and subset $S \subseteq X$, we say that

- x is a **limit point** of S if every open ball centered at x contains a point $y \neq x$ such that $y \in S$;
- S is **closed** if every limit point of S belongs to S;
- x is an **interior point** of S if there is an open ball B centered at x such that $B \subseteq S$;
- S is **open** if every point of S is an interior point of S;
- the **complement** of S (denoted by S^c) is the set of all points $y \in X$ such that $y \notin S$;
- S is **bounded** if there exists an $M < \infty$ and a point $\bar{y} \in X$ such that $d(y, \bar{y}) \leq M$ for all $y \in S$; i.e., $S \subseteq B(\bar{y}, M)$;

- S is **connected** if it cannot be written as the disjoint union of two non–empty open sets.

Given a set $S \subseteq X$, let S' be the set of limit points of S in X. Then, the set $S \cup S'$ is called the **closure** of S and is denoted by $\text{cl}(S)$.

To illustrate the above concepts, consider the case where $S = [0, 1) \subseteq \mathbb{R}$. The point $x = 1$ is a limit point of S, since for any $r > 0$, the open ball $B^\circ(1, r) = \{x \in \mathbb{R} : |x - 1| < r\}$ contains the point $y = 1 - r/2 \neq x$, which belongs to S. However, the limit point $x = 1$ does not belong to S and so S is not closed. Now, any point $\bar{x} \in (0, 1)$ is an interior point of S, since the open ball $B^\circ(\bar{x}, r(\bar{x}))$, where $r(\bar{x}) = \min\{\bar{x}, 1 - \bar{x}\}$, is completely contained in S. On the other hand, $x = 0$ is not an interior point of S and hence S is not open. This example also demonstrates an interesting point: a set can be neither open nor closed! It is clear that S is bounded, as $S \subseteq [-1, 1] = B(0, 1)$. Finally, we have $\text{cl}(S) = [0, 1]$.

Based on the above definitions, we have the following results:

- A set S is open iff its complement is closed.
- A set S is closed iff its complement is open.
- For any collection $\{S_\alpha\}_\alpha$ of open sets, the set $\cup_\alpha S_\alpha$ is open.
- For any collection $\{S_\alpha\}_\alpha$ of closed sets, the set $\cap_\alpha S_\alpha$ is closed.
- For any *finite* collection $S_1, \ldots, S_n$ of open sets, the set $\cap_{i=1}^n S_i$ is open.
- For any *finite* collection $S_1, \ldots, S_n$ of closed sets, the set $\cup_{i=1}^n S_i$ is closed.

Note that the finiteness assumption is crucial for the last two statements to hold. Indeed, consider the collection of sets $S_i = (-1/i, 1/i) \subseteq \mathbb{R}$, where $i = 1, 2, \ldots$. It is clear that each S_i is an open set. However, we have $\cap_{i \geq 1} S_i = \{0\}$, which is closed.

Now, let us turn to the concept of compactness. We say that a subset $S \subseteq \mathbb{R}^n$ is **compact** if it is closed and bounded. One interesting feature of compact sets is the following:

- Every infinite subset S' of a compact set S has a limit point in S.

Moreover,

- Every bounded infinite subset of $\mathbb{R}^n$ has a limit point in $\mathbb{R}^n$.

2 Sequences

Before we discuss sequences in detail, let us introduce various notions of a bound on a set in $\mathbb{R}$. We say that a set $S \subseteq \mathbb{R}$ is **bounded above** if there exists a constant $\beta \in \mathbb{R}$ such that $x \leq \beta$ for all $x \in S$. The number β is called an **upper bound** of S. The **lower bound** of a set in $\mathbb{R}$ is defined analogously.

Now, suppose that $S \subseteq \mathbb{R}$ is bounded above and there exists an $\alpha \in \mathbb{R}$ satisfying (i) α is an upper bound of S and (ii) if $\gamma < \alpha$, then γ is not an upper bound of S. Then, α is called the **least upper bound** or **supremum** of S, which we denote by $\alpha = \sup S$.

In a similar fashion, suppose that $S \subseteq \mathbb{R}$ is bounded below and there exists an $\alpha \in \mathbb{R}$ satisfying (i) α is a lower bound of S and (ii) if $\gamma > \alpha$, then γ is not a lower bound of S. Then, α is called the **greatest lower bound** or **infimum** of S, which we denote by $\alpha = \inf S$. Note that both the supremum and infimum of a set, if exist, are unique.

To illustrate the above concepts, consider the set $S = (0,1) \subseteq \mathbb{R}$. Clearly, S is bounded above by 1 and bounded below by 0. Now, we claim that $1 = \sup S$ and $0 = \inf S$. Indeed, if $\gamma < 1$, then γ cannot be an upper bound of S, for $\gamma < \gamma + \epsilon \in S$ for sufficiently small $\epsilon > 0$. A similar argument shows that $0 = \inf S$, as desired. This example also shows that the supremum and infimum of a set S need not belong to S.

We say that a sequence $\{a_n\}_{n\geq 1}$ in a metric space (X,d) **converges** to $a \in X$ if for every $\epsilon > 0$, there exists an N such that $d(a_n, a) < \epsilon$ whenever $n \geq N$. In this case, we write $a_n \to a$ or $\lim_{n\to\infty} a_n = a$ and call $a \in X$ a **limit** of $\{a_n\}_{n\geq 1}$.

Given a sequence $\{a_n\}_{n\geq 1}$ in (X,d) and a sequence $\{n_j\}_{j\geq 1}$ of positive integers with $n_1 < n_2 < \cdots$, the sequence $\{a_{n_j}\}_{j\geq 1}$ is called a **subsequence** of $\{a_n\}_{n\geq 1}$.

Now, we have the following results concerning sequences and subsequences:

- Let $\{a_n\}_{n\geq 1}, \{b_n\}_{n\geq 1}$ be sequences in $\mathbb{R}^m$ and $\{\alpha_n\}_{n\geq 1}$ be a sequence in $\mathbb{R}$. Suppose that $a_n \to a$, $b_n \to b$, and $\alpha_n \to \alpha$. Then, we have

  $$a_n + b_n \to a + b, \quad a_n b_n \to ab, \quad \alpha_n a_n \to \alpha a,$$

  where $a_n b_n \in \mathbb{R}^m$ denotes the component–wise product of the vectors a_n and b_n.

- If $a, a' \in X$ are such that $a_n \to a$ and $a_n \to a'$, then $a = a'$. In other words, the limit of a sequence, if exists, is unique.

- If $\{a_n\}_{n\geq 1}$ converges, then the set $\{a_n\}_{n\geq 1}$ is bounded.

- The sequence $\{a_n\}_{n\geq 1}$ converges to a iff every subsequence of $\{a_n\}_{n\geq 1}$ converges to a.

- If $\{a_n\}_{n\geq 1}$ is a sequence in a compact set $S \subseteq \mathbb{R}^m$, then there exists a subsequence of $\{a_n\}_{n\geq 1}$ that converges to a point in S.

- Every bounded sequence in $\mathbb{R}^m$ contains a convergent subsequence.

- If $\{a_n\}_{n\geq 1}$ is a **monotonic** sequence in $\mathbb{R}$ (i.e., either $a_n \leq a_{n+1}$ or $a_n \geq a_{n+1}$ for $n = 1, 2, \ldots$), then it converges iff it is bounded.

As an illustration, consider the set $S = [-1, 1] \subseteq \mathbb{R}$. Since S is closed and bounded, it is compact. Now, consider the sequence $a_n = (-1)^n$, where $n = 1, 2, \ldots$. Note that the sequence $\{a_n\}_{n\geq 1}$ does not converge. However, it has two convergence subsequences; namely, $\{a_n\}_{n \text{ odd}}$ and $\{a_n\}_{n \text{ even}}$. The former converges to $-1 \in S$, while the latter converges to $1 \in S$.

Finally, given a sequence $\{a_n\}_{n\geq 1}$ in $\mathbb{R}$, we define its **limit superior** and **limit inferior** by

$$\limsup \{a_n\} = \inf_{n\geq 1} \sup_{j\geq n} \{a_j\}$$

and

$$\liminf \{a_n\} = \sup_{n\geq 1} \inf_{j\geq n} \{a_j\},$$

respectively. We then have the following results:

- $a = \limsup \{a_n\}$ iff (i) for every $\epsilon > 0$, there exists an N such that $a_n < a + \epsilon$ for all $n \geq N$ and (ii) for every $\epsilon > 0$ and integer $N \geq 1$, there exists an $n \geq N$ such that $a_n > a - \epsilon$.

- $a = \liminf \{a_n\}$ iff (i) for every $\epsilon > 0$, there exists an N such that $a_n > a - \epsilon$ for all $n \geq N$ and (ii) for every $\epsilon > 0$ and integer $N \geq 1$, there exists an $n \geq N$ such that $a_n < a + \epsilon$.

- $\liminf \{a_n\} \leq \limsup \{a_n\}$.

- The sequence $\{a_n\}_{n\geq 1}$ converges to a iff $a = \liminf \{a_n\} = \limsup \{a_n\}$.

As an example, consider again the sequence $a_n = (-1)^n$, where $n = 1, 2, \ldots$. For any $n \geq 1$, we have $\sup_{j\geq n} \{a_j\} = 1$ and $\inf_{j\geq n} \{a_j\} = -1$, whence $\limsup \{a_n\} = 1$ and $\liminf \{a_n\} = -1$.

3 Functions

In this section, let (X, d_X), (Y, d_Y) be metric spaces and $S \subseteq X$ be a subset of X. Furthermore, let $f : S \to Y$ be a function.

We begin with the notion of a **limit of a function**. Let $\bar{x} \in X$ be a limit point of S (note that $\bar{x}$ need not be in S). Then, we write $\lim_{x\to\bar{x}} f(x) = \bar{y}$ if there exists a $\bar{y} \in Y$ with the following property: for every $\epsilon > 0$, there exists a $\delta > 0$ such that $d_Y(f(x), \bar{y}) < \epsilon$ for all $x \in S$ with $0 < d(x, \bar{x}) < \delta$.

We can also rephrase the above definition in terms of limits of sequences. Specifically, we have $\lim_{x\to\bar{x}} f(x) = \bar{y}$ iff $\lim_{n\to\infty} f(x_n) = \bar{y}$ for every sequence $\{x_n\}_{n\geq 1}$ in S such that $x_n \neq \bar{x}$ and $\lim_{n\to\infty} x_n = \bar{x}$. From this definition, we see that the limit of a function, if exists, is unique.

3.1 Continuity

Let $x \in S$. We say that f is **continuous** at x if for every $\epsilon > 0$, there exists a $\delta > 0$ such that $d_Y(f(x), f(x')) < \epsilon$ for all $x' \in S$ with $d_X(x, x') < \delta$. If f is continuous at every $x \in S$, then we say that f is continuous on S.

The notion of continuity has many characterizations. We list some of them below.

- Let $\bar{x} \in S$ be a limit point of S. Then, f is continuous at $\bar{x}$ iff $\lim_{x\to\bar{x}} f(x) = f(\bar{x})$.

- f is continuous on S iff the set $f^{-1}(T) \equiv \{x \in S : f(x) \in T\}$ is open for every open set $T \subseteq Y$.

- f is continuous on S iff the set $f^{-1}(T)$ is closed for every closed set $T \subseteq Y$.

Moreover, the following hold:

- Let $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$ be continuous on S. Then, $f + g$ and fg are continuous on S. If $g \neq 0$ on S, then f/g is also continuous on S.

- Let X, Y, Z be metric spaces and $S \subseteq X$ be a subset of X. Consider the functions $f : S \to f(S) \subseteq Y$ and $g : f(S) \to Z$. Suppose that f is continuous at $x \in S$ and g is continuous at $f(x) \in f(S) \subseteq Y$. Then, the function $h : S \to Z$, which is given by $h(x) = g \circ f(x) = g(f(x))$, is continuous at $x \in S$.

- **(Intermediate Value Theorem)** If S is connected in X and f is continuous on S, then $f(S)$ is connected in Y. In particular, if S is connected, $f : S \to \mathbb{R}$ is continuous, and there exist $x, x' \in S$ such that $f(x) < r < f(x')$ for some $r \in \mathbb{R}$, then there exists an $\bar{x} \in S$ such that $f(\bar{x}) = r$.

An important property of continuous functions is that they behave well on compact sets. Before we make this statement precise, let us state a definition. Let X be any set. We say that a function $f : X \to \mathbb{R}^n$ is **bounded** if there exists an $M < \infty$ such that $\|f(x)\|_2 \le M$ for all $x \in X$.

Now, we have the following results concerning continuous functions on compact sets in $\mathbb{R}^l$:

- Let $S \subseteq \mathbb{R}^l$ be a compact set and $f : S \to \mathbb{R}^n$ be a continuous function on S. Then, the set $f(S) \equiv \{y \in \mathbb{R}^n : y = f(x) \text{ for some } x \in S\}$ is closed and bounded. In particular, f is bounded.

- Let $S \subseteq \mathbb{R}^l$ be a compact set and $f : S \to \mathbb{R}$ be a continuous function on S. Define

$$\overline{m} = \sup_{x \in S} f(x) = \sup f(S) \quad \text{and} \quad \underline{m} = \inf_{x \in S} f(x) = \inf f(S).$$

  Then, there exist $x', x'' \in S$ such that $\overline{m} = f(x')$ and $\underline{m} = f(x'')$; i.e., the function f attains its maximum and minimum over S at $x' \in S$ and $x'' \in S$, respectively.

The last result is very important in the context of optimization, as it gives a sufficient condition under which an optimization problem has an optimal solution. As an illustration, let $S = (0, \infty)$ and consider the function $f : S \to \mathbb{R}$ defined by $f(x) = 1/x$. Clearly, S is not compact, as it is neither closed nor bounded. Now, observe that $\inf_{x \in S} f(x) = 0$. However, there does not exist an $x' \in S$ such that $f(x') = 0$.

A stronger notion of continuity is that of **Lipschitz continuity**. Let $S \subseteq X$ and $f : S \to Y$. We say that f is Lipschitz continuous on S with parameter $L \in (0, \infty)$ if for any $x, x' \in S$, we have

$$d_Y(f(x), f(x')) \le L \cdot d_X(x, x').$$

It is clear that f is continuous on S whenever it is Lipschitz continuous on S. However, the converse does not hold in general. For instance, consider the function $f : \mathbb{R} \to \mathbb{R}_+$ defined by $f(x) = |x|^{1/2}$. It can be easily verified that f is continuous but not Lipschitz continuous on $\mathbb{R}$.

3.2 Differentiation

3.2.1 Univariate Functions

In this course, we will frequently need to compute the derivatives of functions. Before we discuss multivariate differentiation, let us first review some of the elements in the theory of univariate differentiation. To begin, let $f : [a, b] \to \mathbb{R}$ be a function. For any $x \in (a, b)$, define

$$\phi(t) = \frac{f(t) - f(x)}{t - x} \quad \text{where } a < t < b,\ t \ne x$$

and set $f'(x) = \lim_{t \to x} \phi(t)$, provided that the limit exists in accordance with the definition given at the beginning of Section 3. The function f' is called the **derivative** of f, and whenever f' is defined at $x \in (a, b)$, we say that f is **differentiable** at x. If f' is defined at every point of a set $S \subseteq (a, b)$, then we say that f is differentiable on S. Note that if f is differentiable at $x \in (a, b)$, then it is continuous at x.

Before we proceed further, let us recall three fundamental results in the theory of differentiation.

- Let $f : [a, b] \to \mathbb{R}$ be continuous. Suppose that $f'(x)$ exists at some $x \in (a, b)$. Furthermore, suppose that g is defined on an interval I that contains the range of f (i.e., $f([a, b]) \subseteq I$) and g is differentiable at $f(x)$. If we define a function $h : [a, b] \to \mathbb{R}$ by $h(t) = g(f(t))$, then the **Chain Rule** asserts that h is differentiable at x and $h'(x) = g'(f(x))f'(x)$.

- Let $f, g : [a, b] \to \mathbb{R}$ be continuous functions that are differentiable on (a, b). Then, the **Mean Value Theorem** asserts the existence of an $x \in (a, b)$ at which

  $$(f(b) - f(a))g'(x) = (g(b) - g(a))f'(x).$$

  By taking $g(x) = x$, we see that if $f : [a, b] \to \mathbb{R}$ is a continuous function that is differentiable on (a, b), then there exists an $x \in (a, b)$ such that

  $$f(b) - f(a) = (b - a)f'(x).$$

- Let $f : [a, b] \to \mathbb{R}$ be continuous and define $F : [a, b] \to \mathbb{R}$ by

  $$F(x) = \int_a^x f(t)\, dt.$$

  Then, the **Fundamental Theorem of Calculus** asserts that F is continuous on $[a, b]$ and differentiable on (a, b). Moreover, we have $F'(x) = f(x)$ for all $x \in (a, b)$.

Now, if f has a derivative f' on an interval and if f' is itself differentiable, then we can speak of the derivative of f', which we denote by f'', and call it the **second derivative** of f. By continuing in this manner, we obtain functions $f, f', f'', f^{(3)}, \ldots, f^{(n)}, \ldots$, each being the derivative of the preceding one. We call $f^{(n)}$ the n**–th derivative** of f.

In many occassions, we will need to approximate a given function by a simpler function, say, a polynomial. One such approximation is given by **Taylor's Theorem**, which asserts the following. Let $f : [a, b] \to \mathbb{R}$ be a function that satisfies the following properties:

(i) $f^{(n-1)}$ is continuous on $[a, b]$.

(ii) $f^{(n)}(t)$ exists for every $t \in (a, b)$.

Let $a \le t_1 < t_2 \le b$ and define

$$P(t) = \sum_{j=0}^{n-1} \frac{f^{(j)}(t_1)}{j!}(t - t_1)^j.$$

Then, there exists a $t_0 \in [t_1, t_2]$ such that

$$f(t_2) = P(t_2) + \frac{f^{(n)}(t_0)}{n!}(t_2 - t_1)^n.$$

In particular, we see that f can be approximated by a polynomial of degree $n - 1$, and the error can be estimated if we have bounds on $|f^{(n)}(x)|$.

3.2.2 Multivariate Functions

Let us now turn our attention to multivariate functions. Let $S \subseteq \mathbb{R}^m$ and $f : S \to \mathbb{R}^n$. Suppose that S contains an open ball centered at $x \in \mathbb{R}^m$. Given $u \in \mathbb{R}^m \backslash \{\mathbf{0}\}$, define

$$f'(x, u) = \lim_{t \searrow 0} \frac{f(x + tu) - f(x)}{t},$$

provided that the limit exists. The limit is called the **directional derivative** of f at x with respect to u. Note that $f'(x, u) \in \mathbb{R}^n$.

As it turns out, the notion of directional derivative is not the most appropriate generalization of the notion of derivative (see, e.g., [2, Chapter 2]). Thus, we need another definition. Towards that end, let $S \subseteq \mathbb{R}^m$ and $f : S \to \mathbb{R}^n$. Suppose that S contains an open ball centered at $x \in \mathbb{R}^m$. We say that f is **differentiable** at x if there exists an $n \times m$ matrix A such that

$$\frac{f(x+u) - f(x) - Au}{\|u\|_2} \to \mathbf{0} \quad \text{as } u \to \mathbf{0}. \tag{1}$$

The matrix A, which is unique, is called the **derivative** of f at x and is denoted by $Df(x)$.

As an example, consider $f : \mathbb{R}^m \to \mathbb{R}^n$ given by $f(x) = Bx + b$, where $B \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$. For any $u \in \mathbb{R}^m$, we have $f(x+u) - f(x) = Bu$. Thus, by taking $A = B$ in (1), we see that $Df(x) = B$.

The following result establishes the relationship between the derivative and directional derivatives of a function:

- Let $S \subseteq \mathbb{R}^m$ and $f : S \to \mathbb{R}^n$. If f is differentiable at x, then all the directional derivatives of f at x exist. Moreover, we have $f'(x,u) = Df(x)u$ for any $u \in \mathbb{R}^m \backslash \{\mathbf{0}\}$.

So far it is a bit awkward to apply the definition directly to compute the derivative of a multivariate function. To simplify the computation, let us introduce another definition. Let $S \subseteq \mathbb{R}^m$ and $f : S \to \mathbb{R}$. We define the i–**th partial derivative** of f at x, denoted by $D_i f(x)$, to be the directional derivative of f at x with respect to the i–th basis vector e_i, provided that it exists. In other words, we define

$$D_i f(x) = \lim_{t\to 0} \frac{f(x + te_i) - f(x)}{t} \qquad \text{for } i = 1, 2, \dots, m.$$

Note that $D_i f(x) \in \mathbb{R}$, as f is a real–valued function. Now, it can be shown that whenever $f : S \to \mathbb{R}$ is differentiable at x, we have

$$Df(x) = \left[\begin{array}{cccc} D_1 f(x) & D_2 f(x) & \cdots & D_m f(x) \end{array}\right] \in \mathbb{R}^m.$$

We remark that the column vector $Df(x)^T$ is also known as the **gradient** of f and is denoted by $\nabla f(x)$.

To generalize the above result to vector–valued functions (i.e., functions of the form $f : S \to \mathbb{R}^n$), we first write $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$, where $f_1, \dots, f_n : S \to \mathbb{R}$. Then, the following hold:

- f is differentiable at x iff each f_i is differentiable at x.
- If f is differentiable at x, then we have

$$Df(x) = \begin{bmatrix} Df_1(x) \\ \vdots \\ Df_n(x) \end{bmatrix} \in \mathbb{R}^{n\times m}.$$

The matrix $Df(x)$ is sometimes known as the **Jacobian matrix** of f.

By viewing Df as a function from $\mathbb{R}^m$ to $\mathbb{R}^{n\times m} \cong \mathbb{R}^{mn}$, we may define the derivative of Df, denoted by $D^2 f$, using the above results, provided that the derivative exists. For instance, when $n = 1$, we have

$$D^2 f(x) = \begin{bmatrix} D_1 D_1 f(x) & D_2 D_1 f(x) & \cdots & D_m D_1 f(x) \\ D_1 D_2 f(x) & D_2 D_2 f(x) & \cdots & D_m D_2 f(x) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 D_m f(x) & D_2 D_m f(x) & \cdots & D_m D_m f(x) \end{bmatrix} \in \mathbb{R}^{m \times m}.$$

In this case, the matrix $(D^2 f(x))^T$ is also known as the **Hessian** of f and is denoted by $\nabla^2 f(x)$.

Finally, we can formulate the **Multivariate Chain Rule** as follows. Let $S \subseteq \mathbb{R}^m$ and $T \subseteq \mathbb{R}^n$. Let $f : S \to \mathbb{R}^n$ and $g : T \to \mathbb{R}^p$ be such that $f(S) \subseteq T$. Let $x \in S$ and suppose that $y = f(x)$. If f is differentiable at x and g is differentiable at y, then the function $h : S \to \mathbb{R}^p$ defined by $h(u) = g(f(u))$ is differentiable at x. Moreover, we have

$$Dh(x) = Dg(f(x))Df(x),$$

where the product on the right–hand side is simply matrix multiplication.

References

[1] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis.* Dover Publications, Inc., New York, 1975.

[2] J. R. Munkres. *Analysis on Manifolds.* Westview Press, Boulder, Colorado, 1991.

[3] H. L. Royden. *Real Analysis.* Macmillan Publishing Company, New York, third edition, 1988.

[4] W. Rudin. *Principles of Mathematical Analysis.* International Series in Pure and Applied Mathematics. McGraw–Hill Book Co., Singapore, third edition, 1976.