Introduction to Analysis II

Kon Yi

February 26, 2026

Abstract

Lecture note of Introduction to Analysis II.

Contents

Chapter 1

Several Variable Differential Calculus

Lecture 1

In this chapter, we want to approximate non-linear functions by linear maps. If we consider 24 Feb.

$$f : \mathbb{R}^n \to \mathbb{R}^m \quad f(\underbrace{x_1, x_2, \dots, x_n}_{x}) = (f_1(x), f_2(x), \dots, f_m(x)),$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$ for all i. Now given a real-valued function $F : \mathbb{R}^n \to \mathbb{R}$ which is differentiable at a point x_0. We know that near $x_0 \in \mathbb{R}^n$ we can approximate $F(x)$ in the following way:

$$F(x) \approx F(x_0) + \nabla F(x_0) \cdot (x - x_0)$$

where

$$\nabla F(x_0) = \left(\frac{\partial F(x_0)}{\partial x_1}, \frac{\partial F(x_0)}{\partial x_2}, \dots, \frac{\partial F(x_0)}{\partial x_n} \right) \in \mathbb{R}^n \text{ with } x_0 = (x_1, x_2, \dots, x_n)$$

and thus

$$\begin{aligned} \nabla F(x_0) \cdot (x - x_0) &= \left\langle \frac{\partial F(x_0)}{\partial x_1}, \frac{\partial F(x_0)}{\partial x_2}, \dots, \frac{\partial F(x_0)}{\partial x_n} \right\rangle \cdot \langle x_1, x_2, \dots, x_n \rangle \\ &= \sum_{i=1}^{n} \frac{\partial F(x_0)}{\partial x_i} x_i. \end{aligned}$$

Hence,

$$f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{pmatrix} \approx \begin{pmatrix} f_1(x_0) + \nabla f_1(x)(x - x_0) \\ f_2(x_0) + \nabla f_2(x)(x - x_0) \\ \vdots \\ f_n(x_0) + \nabla f_n(x)(x - x_0) \end{pmatrix},$$

which gives

$$f(x) - f(x_0) \approx \begin{pmatrix} \nabla f_1(x)(x - x_0) \\ \nabla f_2(x)(x - x_0) \\ \vdots \\ \nabla f_n(x)(x - x_0) \end{pmatrix} = \begin{pmatrix} \nabla f_1(x) \\ \nabla f_2(x) \\ \vdots \\ \nabla f_n(x) \end{pmatrix} \cdot \underbrace{(x - x_0)}_{\text{column vector}}.$$

1.1 Linear Transformation

Definition 1.1.1 (Row vectors). Let $n \geq 1$ be an integer. We refer to elements of $\mathbb{R}^n$ an n-dimensional row vectors. A typical row vector is $x = (x_1, x_2, \ldots, x_n)$ which we abbreviate as $(x_i)_{1 \leq i \leq m}$. The components $x_1, x_2, \ldots, x_n$ are real numbers. If x and y are two row vectors in $\mathbb{R}^n$, we can define vector sum by

$$x + y = (x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) := (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

If $c \in \mathbb{R}$ is any real number, we define scalar multiplications by

$$cx = c(x_1, x_2, \ldots, x_n) = (cx_1, cx_2, \ldots, cx_n).$$

Remark 1.1.1.

(1) $-x := (-1) \cdot x = (-x_1, -x_2, \ldots, -x_n)$.

(2) zero vector is denoted by 0, i.e. $(0, 0, \ldots, 0)$.

Lemma 1.1.1 ($\mathbb{R}^n$ is a vector space). Let x, y, z be vectors in $\mathbb{R}^n$, and let $c, d \in \mathbb{R}$. Then the following properties hold:

(a) $x + y = y + x$.

(b) $(x + y) + z = x + (y + z)$.

(c) $x + 0 = 0 + x = x$.

(d) $x + (-x) = (-x) + x = 0$.

(e) $(c \cdot d)x = c \cdot (dx)$.

(f) $c(x + y) = cx + cy$.

(g) $(c + d)x = cx + dx$.

(h) $1x = x$.

Definition 1.1.2. Let $x = (x_1, x_2, \ldots, x_n)$ be row vector. Its transpose is the n-dimensional column vector

$$x^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Definition 1.1.3. The standard basis of $\mathbb{R}^n$ consists of $e_1, e_2, \ldots, e_n$, where e_j has 1 in the j-th position and 0 elsewhere:

$$e_j = (0, \ldots, 0, \underbrace{1}_{j\text{-th}}, 0, \ldots, 0).$$

Every row vector

$$x = (x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} x_j e_j.$$

Similarly,

$$x^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{j=1}^{n} x_j e_j^T.$$

Definition 1.1.4 (Linear transformation)**.** A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is any function from one Euclidean space to another that satisfies the following two properties:

(a) Additivity: For $x, y \in \mathbb{R}^n$, $T(x + y) = T(x) + T(y)$.

(b) Homogenity: For $x \in \mathbb{R}^n$ and all scalars $c \in \mathbb{R}$, $T(cx) = cT(x)$.

Remark 1.1.2. This definition is equivalent to the following:

$$T(c_1v_1 + \cdots + c_kv_k) = c_1T(v_1) + \cdots + c_kT(v_k)$$

where $v_1, \ldots, v_k \in \mathbb{R}^n$ and $c_1, c_2, \ldots, c_n \in \mathbb{R}$.

Definition 1.1.5. Let $m, n \geq 1$ be integers. An $m \times n$ ordered matrix is an ordered rectangular array of real numbers

$$A = (a_{ij})_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

consisting of m rows and n columns, where

(a) The entry a_{ij} denote the number in the i-th row and j-th column.

(b) We denote the set of all $m \times n$ matrices by $\mathbb{R}^{m\times n}$.

(c) In particular, a row vector is a $1 \times n$ matrix, a column vector is a $n \times 1$ vector.

Definition 1.1.6 (Matrix multiplication)**.** Given an $m \times n$ matrix $A = (a_{ij})_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$ and an $n \times p$ matrix $B = (b_{jk})_{\substack{1\leq j\leq n \\ 1\leq k\leq p}}$, we define AB to be the $m \times p$ matrix $(c_{ik})_{\substack{1\leq i\leq m \\ 1\leq k\leq p}}$ where

$$c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk}.$$

Definition 1.1.7 (Matrix-vector multiplication)**.** Let $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ and $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ be a column vector. We define

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{pmatrix}$$

Remark 1.1.3. In our class, we just treat $\mathbb{R}^n, \mathbb{R}^m$ as column vector spaces, and $L_A(x) = Ax$ is a $m \times 1$ column vector.

Theorem 1.1.1. Let A be a $m \times n$ matrix, then $L_A(x) = Ax$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Proof. ∎ DIY

Proposition 1.1.1. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. For each $j = 1, 2, \ldots, n$, let e_j denote the j-th standard basis vector in $\mathbb{R}^n$ and write $T(e_j) = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$. Define the matrix $A = (a_{ij})$, then $T(x) = Ax$.

Proof. Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. We can write $x = \sum_{j=1}^n x_j e_j$, then we know

$$T(x) = T\left(\sum_{j=1}^n x_j e_j\right) = \sum_{j=1}^n x_j T(e_j) = \sum_{j=1}^n x_j \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = Ax.$$

■

Lemma 1.1.2. Let A be a $m \times n$ matrix and let B be a $n \times p$ matrix. Then $L_A \circ L_B = L_(AB)$.

Proof. It suffices to show that $(L_A \circ L_B)(x) = L_{AB}(x)$ for $x \in \mathbb{R}^p$, and the rest is easy. ■

As previously seen. $f : E \to \mathbb{R}$ where E is a subset of $\mathbb{R}$, then

$$f'(x_0) = \lim_{\substack{x \to x_0 \\ x \in E \setminus \{x_0\}}} \frac{f(x) - f(x_0)}{x - x_0}.$$

Suppose now $f : E \subseteq \mathbb{R}^n \to \mathbb{R}^m$, we can't define

$$f'(x) = \lim_{\substack{x \to x_0 \\ x \in E \setminus \{x_0\}}} \frac{f(x) - f(x_0)}{x - x_0}$$

since the denominator and the numerator are vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$.

1.2 Derivatives in Several Variable Calculus

Lemma 1.2.1. Let $E \subseteq \mathbb{R}$, let $f : E \to \mathbb{R}$ be a function and let $L \in \mathbb{R}$ and x_0 be a limit point of E. Then the following two statements are equivalent:

(a) f is differentiable at x_0 and $f'(x_0) = L$.

(b) $\lim_{\substack{x \to x_0 \\ x \in E \setminus \{x_0\}}} \frac{|f(x) - f(x_0) - L(x - x_0)|}{|x - x_0|} = 0$.

Proof. Note that

$$\frac{f(x) - f(x_0)}{x - x_0} = L + \frac{f(x) - f(x_0) - L(x - x_0)}{x - x_0} \text{ if } x \neq x_0,$$

so we have

$$\frac{f(x) - f(x_0)}{x - x_0} - L = \frac{f(x) - f(x_0) - L(x - x_0)}{x - x_0} \text{ if } x \neq x_0,$$

and thus

$$0 = \lim_{\substack{x \to x_0 \\ x \in E \setminus \{x_0\}}} \left| \frac{f(x) - f(x_0)}{x - x_0} - L \right| = \lim_{\substack{x \to x_0 \\ x \in E \setminus \{x_0\}}} \left| \frac{f(x) - f(x_0) - L(x - x_0)}{x - x_0} \right|.$$

■

Lecture 2

26 Feb.

Definition 1.2.1 (Differentiability). Let E be a subset of $\mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a function, and let x_0 be a limit point of E. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say f is differentiable at x_0 with derivative L if

$$\lim_{\substack{x \to x_0 \\ x \in E \setminus \{x_0\}}} \frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} = 0,$$

or equivalently, given $\varepsilon > 0$, $\exists \delta > 0$ s.t. for all $x \in E$ satisfying $0 < \|x - x_0\| < \delta$, we have

$$\frac{\|f(x) - f(x_0) - L(x - x_0)\|}{\|x - x_0\|} < \varepsilon.$$

Remark 1.2.1. $\|f(x) - f(x_0) - L(x - x_0)\|$ is the length of a vector in $\mathbb{R}^m$ and $\|x - x_0\|$ is the length of a vector in $\mathbb{R}^n$.

Remark 1.2.2. x_0 is a limit point of $E \subseteq \mathbb{R}^n$ if for any $r > 0$, $B(x_0, r) \cap (E \setminus \{x_0\}) \neq \varnothing$. That is, for every $r > 0$, $\exists x \in E$ s.t. $0 < \|x - x_0\| < r$.

Example 1.2.1. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ has

$$f(x) = \begin{cases} 0, & \text{if } x \in E = \{(x_1, x_2, \ldots, x_n) \mid x_n \geq 0\}; \\ x, & \text{if } x \in \mathbb{R}^n \setminus E \end{cases}$$

then f is differentiable on E and $f'(x) = 0$ and f is not differentiable on $\mathbb{R}^n$ at 0.

Remark 1.2.3. Recall that x_0 is said to be an interior point of $E \subseteq \mathbb{R}^n$ if there exists $r > 0$ s.t.

$$B(x_0, r) = \{x \in \mathbb{R}^n : \|x - x_0\| < r\} \subseteq E.$$

If x_0 is an interior point of E, then f is differentiable at x_0 is equivalent to

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - L(h)\|}{\|h\|} = 0$$

since if $\|h\| < r$, then $x_0 + h \in B(x_0, r)$, which gives $x_0 + h \in E$, and thus $f(x_0 + h)$ is well-defined.

Remark 1.2.4. Here $\|\cdot\|$ denoted the standard Euclidean norm on $\mathbb{R}^n$ (and on $\mathbb{R}^m$):

$$\|(x_1, x_2, \ldots, x_n)\| = \left(x_1^2 + x_2^2 + \cdots + x_n^2\right)^{\frac{1}{2}}.$$

Example 1.2.2. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f(x, y) = \left(x^2, y^2\right),$$

let $x_0 = (1, 2)$, and let $L : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map

$$L(x, y) = (2x, 4y).$$

We claim that f is differentiable at x_0 with derivative L.

Proof. By definition, f is differentiable at x_0 with derivative L if and only if

$$\lim_{h\to 0} \frac{\|f(x_0+h)-f(x_0)-L(h)\|}{\|h\|} = 0,$$

where $h=(a,b)$. Thus, we know the above equation becomes

$$\frac{\|f(1+a,2+b)-f(1,2)-L(a,b)\|}{\|(a,b)\|} = \frac{\sqrt{a^4+b^4}}{\sqrt{a^2+b^2}},$$

and since

$$0 \leq \frac{\sqrt{a^4+b^4}}{\sqrt{a^2+b^2}} \leq \sqrt{a^2+b^2},$$

so by the Squeeze Theorem, we have

$$\lim_{(a,b)\to(0,0)} \frac{\sqrt{a^4+b^4}}{\sqrt{a^2+b^2}} = 0,$$

and we're done. ⊛

Lemma 1.2.2 (Uniqueness of derivative)**.** Let E be a subset of $\mathbb{R}^n$, and let $f: E\to\mathbb{R}^n$ be a function, and let $x_0\in E$ be an interior point of E. Suppose $L_1, L_2 : \mathbb{R}^n\to\mathbb{R}^m$ are linear transformations s.t. f is differentiable at x_0 with derivative L_1 and also differentiable at x_0 with derivative L_2. Then $L_1 = L_2$.

Proof. Since x_0 is an interior point. By Remark 1.2.3, we have

$$\lim_{h\to 0} \frac{\|f(x_0+h)-f(x_0)-L_1(h)\|}{\|h\|} = 0 \text{ and } \lim_{h\to 0} \frac{\|f(x_0+h)-f(x_0)-L_2(h)\|}{\|h\|} = 0.$$

Thus,

$$\begin{aligned} 0 \leq \frac{\|L_1(h)-L_2(h)\|}{\|h\|} &= \frac{\|L_1(h)-f(x_0+h)+f(x_0)+f(x_0+h)-f(x_0)-L_2(h)\|}{\|h\|} \\ &\leq \frac{\|L_1(h)-f(x_0+h)+f(x_0)\|}{\|h\|} + \frac{\|f(x_0+h)+f(x_0)-L_2(h)\|}{\|h\|}. \end{aligned}$$

Thus, by squeeze theorem, we know

$$\lim_{h\to 0} \frac{\|L_1(h)-L_2(h)\|}{\|h\|} = 0.$$

Now take $v\in\mathbb{R}^n\setminus\{0\}$, and let $h = tv$, so

$$\lim_{t\to 0} \frac{\|L_1(tv)-L_2(tv)\|}{\|tv\|} = 0.$$

That is,

$$\lim_{t\to 0} \frac{|t|\,\|L_1(v)-L_2(v)\|}{|t|\|v\|} = 0.$$

Hence, we have

$$\lim_{t\to 0} \frac{\|L_1(v)-L_2(v)\|}{\|v\|} = 0,$$

and since $v\neq 0$, so we know $L_1(v) = L_2(v)$, and since v can be arbitrary non-zero vector in $\mathbb{R}^n\setminus\{0\}$, so $L_1 = L_2$. $(L_1(0) = L_2(0) = 0)$ ■

Remark 1.2.5. Because of Lemma 1.2.2, the derivative of f at interior points x_0 is unique, and thus we may safely write it as $f'(x_0)$ or $Df(x_0)$. Thus $f'(x_0)$ is the unique linear transformation $f' : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to x_0 \ x \neq x_0} \frac{\|f(x) - (f(x_0) + f'(x_0)(x - x_0))\|}{\|x - x_0\|} = 0.$$

Informally, this condition means that near x_0, the function f can be approximated by its linearization:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

This approximation is sometimes referred to as Newton's approximation in higher dimension.

Remark 1.2.6. A useful consequence of Lemma 1.2.2 is the following: If two functions f and g satisfy $f(x) = g(x)$ for all $x \in E$, and both are differentiable at an interior point x_0, then their derivatives coincide at x_0, i.e. $f'(x_0) = g'(x_0)$. This will be important in later arguments, where one extends functions to larger domains or modifies them on sets of measure zero.

Remark 1.2.7. As we have emphasized, Lemma 1.2.2 guarantees the uniqueness of the derivative only at interior points of the domain E. If x_0 is instead a boundary point of E, the derivative may fail to be uniquely determined.

1.3 Partial and Directional Derivatives

We now begin to relate the concept of differentiability to the more classical notions of partial and directional derivatives. Directional derivatives describe how f changes when we move from a point x_0 in a fixed direction v.

Definition 1.3.1 (Directional derivative). Let $E \subseteq \mathbb{R}^n$, let $f : E \to \mathbb{R}^m$ be a function, let x_0 be an interior point of E and let $v \in \mathbb{R}^n$. If the limit

$$\lim_{\substack{t \to 0, t > 0 \\ x_0 + tv \in E}} \frac{f(x_0 + tv) - f(x_0)}{t} = \lim_{\substack{t \to 0^+ \\ x_0 + tv \in E}} \frac{f(x_0 + tv) - f(x_0)}{t}$$

exists, then we say f is differentiable in the direction v at x_0. This limit is called the directional derivative of f at x_0 in the direction v, and we denote it by

$$D_v f(x_0) \coloneqq \lim_{t \to 0^+} \frac{f(x_0 + tv) - f(x_0)}{t} \in \mathbb{R}^m.$$

Remark 1.3.1. Under this definition,

$$D_{-v} f(x_0) = \lim_{t \to 0^+} \frac{f(x_0 + t(-v)) - f(x_0)}{t} = \lim_{t \to 0^+} \frac{f(x_0 - tv) - f(x_0)}{t}.$$

In usual definition, one define

$$D_v f(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}.$$

This usual definition implies that

$$D_v f(x_0) = -D_{-v} f(x_0)$$

since

$$\begin{aligned} D_v f(x_0) &= \lim_{t\to 0^+} \frac{f(x_0+tv)-f(x_0)}{t} = \lim_{t\to 0^-} \frac{f(x_0+tv)-f(x_0)}{t} \\ &= \lim_{s\to 0^+} = \frac{f(x_0+(-s)v)-f(x_0)}{(-s)} = -\lim_{s\to 0^+} \frac{f(x_0-sv)-f(x_0)}{s} = -(D_{-v}f(x_0)). \end{aligned}$$

Remark 1.3.2. This definition should be compared with the definition of differentiability. Here we divide by the scalar t rather than by a vector, so the expression always makes sense algebraically. The directional derivative $D_v f(x_0)$ is a vector in $\mathbb{R}^m$.

Remark 1.3.3. It is sometimes possible to define directional derivatives at boundary points of E, provided that the vector v points inward toward the domain. However, we shall restrict attention to interior points in what follows.

Lemma 1.3.1. Let $E \subseteq \mathbb{R}^n$, let $f : E \to \mathbb{R}^m$, let $x_0 \in \mathrm{Int}(E)$ and $v \in \mathbb{R}^n$. If f is differentiable at x_0, then f is differentiable in the direction v at x_0, and

$$D_v f(x_0) = f'(x_0)(v).$$

Proof. Since $x_0 \in \mathrm{Int}(E)$ and f is differentiable at x_0, so

$$\lim_{h\to 0} \frac{\|f(x_0+h)-f(x_0)-f'(x_0)(h)\|}{\|h\|} = 0.$$

Note that for $t \neq 0$, we have

$$\frac{f(x_0+tv)-f(x_0)}{t} = \frac{f(x_0+tv)-f(x_0)-f'(x_0)(tv)}{t} + f'(x_0)(v),$$

so

$$\frac{f(x_0+tv)-f(x_0)}{t} - f'(x_0)(v) = \frac{f(x_0+tv)-f(x_0)-f'(x_0)(tv)}{t}.$$

Now we show that

$$\lim_{t\to 0^+} \frac{f(x_0+tv)-f(x_0)-f'(x_0)(tv)}{t} = 0.$$

Recall that f is differentiable at x_0 and $x_0 \in \mathrm{Int}(E)$, so

$$\lim_{h\to 0} \frac{\|f(x_0+h)-f(x_0)-f'(x_0)(h)\|}{\|h\|} = 0.$$

Let $h = tv$ when $v \neq 0$, then

$$\lim_{t\to 0^+} \frac{\|f(x_0+tv)-f(x_0)-f'(x_0)(tv)\|}{\|tv\|} = 0 \Rightarrow \lim_{t\to 0^+} \frac{\|f(x_0+tv)-f(x_0)-f'(x_0)(tv)\|}{|t|} = 0,$$

which is equivalent to

$$\lim_{t\to 0^+} \frac{f(x_0+tv)-f(x_0)-f'(x_0)(tv)}{t} = 0,$$

(since the numerator should tend to 0 so the limit will exist), so we know

$$\lim_{t\to 0^+} \frac{f(x_0+tv)-f(x_0)}{t} - f'(x_0)(v) = 0,$$

i.e.

$$\lim_{t\to 0^+} \frac{f(x_0+tv)-f(x_0)}{t} = f'(x_0)(v).$$

■

Remark 1.3.4. One consequence of this lemma is that total differentiability implies directional differentiability. However, the converse is not true: the existence of all directional derivatives does not guarantee differentiability.

Definition 1.3.2 (Partial Derivative)**.** Let $E \subseteq \mathbb{R}^n$, let $f : E \to \mathbb{R}^m$, and let $x_0 \in \text{Int}(E)$, and let $1 \le j \le n$. The partial derivative of f with respect to the variable x_j at x_0, denoted $\frac{\partial f}{\partial x_j}(x_0)$, is defined by

$$\frac{\partial f}{\partial x_j}(x_0) := \lim_{t \to 0, t \neq 0} \frac{f(x_0 + te_j) - f(x_0)}{t}$$

provided the limit exists.

Remark 1.3.5. If $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$, then differentiation is componentwise:

$$\frac{\partial f}{\partial x_j}(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(x_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x_0) \end{pmatrix}.$$

We say that f is continuously differentiable if all partial derivatives exist and are continuous.

Remark 1.3.6. If f is differentiable at x_0, then

$$\frac{\partial f(x_0)}{\partial x_j} = D_{e_j} f(x_0) = f'(x_0)(e_j).$$

Remark 1.3.7. If f is differentiable at x_0, then

$$D_v f(x_0) = f'(x_0)(v) = f'(x_0)\left(\sum_{j=1}^n v_j e_j\right) = \sum_{j=1}^n v_j f'(x_0)(e_j) = \sum_{j=1}^n v_j \frac{\partial f(x_0)}{\partial x_j}.$$

Theorem 1.3.1. Let $E \subseteq \mathbb{R}^n$, let $f : E \to \mathbb{R}^m$, $F \subseteq E$, and let $x_0 \in \text{Int}(F)$. If all partial derivative $\frac{\partial f}{\partial x_j}$ exists on F and are continuous at x_0, then f is differentiable at x_0, and

$$f'(x_0)(v) = \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x_0)$$

for all $v \in \mathbb{R}^n$.

Proof. Define a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ by

$$L(v) := \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(x_0).$$

We prove that f is differentiable at x_0 with derivative L. It suffices to show that for any $\varepsilon > 0$, $\exists \delta > 0$ s.t.

$$\|f(x) - f(x_0) - L(x - x_0)\| < \varepsilon \|x - x_0\| \text{ whenever } x \in F \text{ and } \|x - x_0\| < \delta$$

Since $x_0 \in \text{Int}(F)$, $\exists r > 0$ s.t. $B(x_0, r) \subseteq F$. Because each partial derivative $\frac{\partial f_i}{\partial x_j}$ is continuous at

x_0, for every pair (i, j) there exists $\delta_{ij} > 0$ s.t.

$$\left|\frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(x_0)\right| < \frac{\varepsilon}{nm} \text{ whenever } \|x - x_0\| < \delta_{ij}.$$

Let $\delta := \min\{r, \delta_{ij} : 1 \le i \le m, 1 \le j \le n\}$. Now fix $x \in F$ with $\|x - x_0\| < \delta$ and write

$$x - x_0 = \sum_{j=1}^{n} v_j e_j, \text{ so } x = x_0 + \sum_{j=1}^{n} v_j e_j.$$

Then

$$\|x - x_0\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad |v_j| \le \|x - x_0\|,$$

Write $f = (f_1, f_2, \ldots, f_m)$. To estimate $f(x) - f(x_0)$, we vary the coordinates one at a time. Consider

$$x^{(0)} = x_0,$$
$$x^{(j)} = x_0 + \sum_{k=1}^{j} v_k e_k \text{ for } j = 1, 2, \ldots, n.$$

Thus, $x^{(n)} = x$ and

$$f(x) - f(x_0) = \sum_{j=1}^{n} \left(f\left(x^{(j)}\right) - f\left(x^{(j-1)}\right)\right) = \sum_{j=1}^{n}\sum_{i=1}^{m} \left(f_i\left(x^{(j)}\right) - f_i\left(x^{(j-1)}\right)\right) e_i.$$

Fix j and consider the j-th increment. For each component f_i, define

$$\varphi_i^j(t) = f_i\left(x^{(j-1)} + t v_j e_j\right), \quad 0 \le t \le 1.$$

Then

$$f_i\left(x^{(j)}\right) - f_i\left(x^{(j-1)}\right) = \varphi_i^j(1) - \varphi_i^j(0).$$

Since f_i has continuous partial derivatives, φ_i^j is differentiable. By MVT, there exists $t_j \in (0, 1)$ s.t.

$$\varphi_i^j(1) - \varphi_i^j(0) = \varphi_i^{j\prime}(t_j).$$

By the chain rule,

$$\varphi_i^{j\prime}(t) = \nabla f_i\left(x^{(j-1)} + t v_j e_j\right) \cdot (v_j e_j) = \frac{\partial f_i}{\partial x_j}\left(x^{(j-1)} + t v_j e_j\right) v_j.$$

Hence,

$$f_i\left(x^{(j)}\right) - f_i\left(x^{(j-1)}\right) = \frac{\partial f_i}{\partial x_j}(\xi_j) v_j,$$

where

$$\xi_j = x^{(j-1)} + t_j v_j e_j.$$

Subtracting $\frac{\partial f_i}{\partial x_j}(x_0) v_j$, we obtain

$$\left|f_i\left(x^{(j)}\right) - f_i\left(x^{(j-1)}\right) - \frac{\partial f_i}{\partial x_j}(x_0) v_j\right| = \left|\frac{\partial f_i}{\partial x_j}(\xi_j) - \frac{\partial f_i}{\partial x_j}(x_0)\right| |v_j|.$$

Because $\|\xi_j - x_0\| < \delta$, continuity of partial derivatives yields,

$$\left|f_i\left(x^{(j)}\right) - f_i\left(x^{(j-1)}\right) - \frac{\partial f_i}{\partial x_j}(x_0) v_j\right| < \frac{\varepsilon}{nm}|v_j|.$$

Summing over $i = 1, 2, \ldots, m$ and using $\|u\| \le \sum_i |u_i|$, we conclude

$$\left\| f\left(x^{(j)}\right) - f\left(x^{(j-1)}\right) - \frac{\partial f}{\partial x_j}(x_0) v_j \right\| = \left\| \sum_{i=1}^{m} \left(f\left(x^{(j)}\right) - f\left(x^{(j-1)}\right) - \frac{\partial f}{\partial x_j}(x_0) v_j \right) e_i \right\|$$

$$\le \sum_{i=1}^{m} \left| f\left(x^{(j)}\right) - f\left(x^{(j-1)}\right) - \frac{\partial f}{\partial x_j}(x_0) v_j \right| \le m \frac{\varepsilon}{nm} |v_j| = \frac{\varepsilon}{n} |v_j| \le \frac{\varepsilon}{n} \|x - x_0\|.$$

Finally, summing over $j = 1, 2, \ldots, n$ and applying the triangle inequality,

$$\left\| f(x) - f(x_0) - \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x_0) v_j \right\| = \left\| \sum_{j=1}^{n} \left(f\left(x^{(j)}\right) - f\left(x^{(j-1)}\right) - \frac{\partial f}{\partial x_j}(x_0) v_j \right) \right\|$$

$$\le \sum_{j=1}^{n} \left\| \left(f\left(x^{(j)}\right) - f\left(x^{(j-1)}\right) - \frac{\partial f}{\partial x_j}(x_0) v_j \right) \right\| \le \sum_{j=1}^{n} \frac{\varepsilon}{n} \|x - x_0\| = \varepsilon \|x - x_0\|.$$

Since $L(x - x_0) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0)$, so we have

$$\|f(x) - f(x_0) - L(x - x_0)\| \le \varepsilon \|x - x_0\|,$$

and we're done. ■

Remark 1.3.8. From Theorem 1.3.1 and Lemma 1.3.1 we conclude the following important fact:

If the partial derivatives of a function $f : E \to \mathbb{R}^m$ exist and are continuous on a set $F \subseteq E$, then at every interior point x_0 of F all directional derivatives exist, and they are given by

$$D_{(v_1, v_2, \ldots, v_n)} f(x_0) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$

- The scalar-valued case. If $f : E \to \mathbb{R}$ is real-valued, we define the gradient of f at x_0 to be the row vector
$$\nabla f(x_0) := \left(\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0) \right).$$
Whenever x_0 lies in the interior of a region where the partial derivatives exist and are continuous, the directional derivative takes the familiar form
$$D_v f(x_0) = v \cdot \nabla f(x_0).$$

- The vector-valued case. Now let $f : E \to \mathbb{R}^m$ be vector-valued, say
$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}.$$
If x_0 lies in the interior of a region where all partial derivatives exist and are continuous, then
$$f'(x_0)(v_1, v_2, \ldots, v_n) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(x_0).$$
Writing this componentwise,
$$f'(x_0)(v_1, \ldots, v_n) = \sum_{j=1}^{n} v_j \begin{pmatrix} \frac{\partial f_1}{\partial x_j}(x_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(x_0) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} v_j \frac{\partial f_1}{\partial x_j}(x_0) \\ \vdots \\ \sum_{j=1}^{n} v_j \frac{\partial f_m}{\partial x_j}(x_0) \end{pmatrix} = \begin{pmatrix} \nabla f_1(x_0) \cdot v \\ \vdots \\ \nabla f_m(x_0) \cdot v \end{pmatrix} = \begin{pmatrix} \nabla f_1(x_0) \\ \vdots \\ \nabla f_m(x_0) \end{pmatrix} v$$

- The derivative matrix (Jacobian matrix). We therefore define the derivative matrix (or Jacobian matrix) of f at x_0 by
$$Df(x_0) \coloneqq \left(\frac{\partial f_i}{\partial x_j}(x_0)\right)_{\substack{1\le i\le m\\ 1\le j\le n}}.$$
Explicitly,
$$Df(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \cdots & \frac{\partial f_m}{\partial x_n}(x_0) \end{pmatrix}.$$
With this notation, the derivative acts by matrix multiplication:
$$D_v f(x_0) = f'(x_0)v = Df(x_0)v.$$

Appendix