Chapter 2

Differentiation of Functions of Several Variables

2.1 Functions of Several Variables

Definition 2.1. Let $\mathcal{V}$ be a vector space (over a scalar field $\mathbb{F}$). A ***$\mathcal{V}$-valued function*** f of n real variables is a rule that assigns a unique vector $f(x_1, \cdots, x_n) \in \mathcal{V}$ to each point $(x_1, \cdots, x_n)$ in some subset A of $\mathbb{R}^n$. The set A is called the ***domain*** of f, and usually is denoted by $\mathrm{Dom}(f)$. The set of vectors $f(x_1, \cdots, x_n)$ obtained from points in the domain is called the ***range*** of f and is denoted by $\mathrm{Ran}(f)$. We write $f : A \to \mathcal{V}$ if f is a $\mathcal{V}$-valued function defined on $A \subseteq \mathbb{R}^n$.

If $\mathcal{V} = \mathbb{R}$, we simply call $f : \mathrm{Dom}(f) \to \mathbb{R}$ a ***real-valued function***, while if $\mathcal{V} = \mathbb{R}^m$, we simply call $f : \mathrm{Dom}(f) \to \mathcal{V}$ as a ***vector-valued function***.

A ***vector field*** is a vector-valued function $f : \mathrm{Dom}(f) \to \mathcal{V}$ such that $\mathrm{Dom}(f) \subseteq \mathcal{V} = \mathbb{R}^n$ for some $n \in \mathbb{N}$.

Definition 2.2. Let $\mathcal{V}$ be a vector space over $\mathbb{R}$, $A \subseteq \mathbb{R}^n$ be a set, and $f, g : A \to \mathcal{V}$ be $\mathcal{V}$-valued functions, $h : A \to \mathbb{R}$ be a real-valued function. The functions $f + g$, $f - g$ and hf, mapping from A to $\mathcal{V}$, are defined by

$$\begin{aligned}(f+g)(x) &= f(x) + g(x) \qquad \forall\, x \in A\,,\\ (f-g)(x) &= f(x) - g(x) \qquad \forall\, x \in A\,,\\ (hf)(x) &= h(x)f(x) \qquad \forall\, x \in A\,.\end{aligned}$$

The map $\dfrac{f}{h} : A \backslash \{x \in A \,|\, h(x) = 0\} \to \mathcal{V}$ is defined by

$$\Big(\frac{f}{h}\Big)(x) = \frac{f(x)}{h(x)} \qquad \forall\, x \in A \backslash \{x \in A \,|\, h(x) = 0\}\,.$$

Definition 2.3. A set $\mathcal{U} \subseteq \mathbb{R}^n$ is said to be open in $\mathbb{R}^n$ if for each $x \in \mathcal{U}$, there exists $r > 0$ such that $B(x, r)$, the ball centered at x with radius r given by

$$B(x, r) = \big\{ y \in \mathbb{R}^n \,\big|\, \|x - y\|_{\mathbb{R}^n} < r \big\}\,,$$

is contained in $\mathcal{U}$. A set $\mathcal{F} \subseteq \mathbb{R}^n$ is said to be closed in $\mathbb{R}^n$ if $\mathcal{F}^{\complement}$, the complement of $\mathcal{F}$, is open in $\mathbb{R}^n$.

Let $A \subseteq \mathbb{R}^n$ be a set. A point x_0 is said to be

1. an ***interior point*** of A if there exists $r > 0$ such that $B(x_0, r) \subseteq A$;

2. an ***isolated point*** of A if there exists $r > 0$ such that $B(x_0, r) \cap A = \{x_0\}$;

3. an ***exterior point*** of A if there exists $r > 0$ such that $B(x_0, r) \subseteq A^{\complement}$;

4. a ***boundary point*** of A if for each $r > 0$, $B(x_0, r) \cap A \neq \varnothing$ and $B(x_0, r) \cap A^{\complement} \neq \varnothing$.

The collection of all interior points of A is called the interior of A and is denoted by $\mathring{A}$. The collection of all exterior points of A is called the exterior of A, and the collection of all boundary point of A is called the boundary of A. The boundary of A is denoted by ∂A. The closure of A is defined as $A \cup \partial A$ and is denoted by $\overline{A}$. The derived set of A, denoted by A', is the collection of all points in $\overline{A}$ that are not isolated points.

A is said to be bounded in $\mathbb{R}^n$ if there exists a constant $M > 0$ such that

$$\|x\|_{\mathbb{R}^n} < M \quad \forall\, x \in A \qquad \big(\Leftrightarrow A \subseteq B(0, M) \big)\,.$$

A is said to be unbounded if A is not bounded.

The following theorem is a fundamental result in point-set topology. We omit the proof since it is not the main concern in vector analysis; however, the result should look intuitive and the proof of this theorem is not difficult. Interested readers can try to establish this result by yourselves.

Theorem 2.4. *Let $A \subseteq \mathbb{R}^n$ be a set. Then*

1. *A is open if and only if* $A = \mathring{A}$;

2. *A is closed if and only if* $A = \overline{A}$;

3. *A is closed if and only if* $\partial A \subseteq A$.

Definition 2.5 (Level Sets, and Graphs)**.** Let $A \to \mathbb{R}^{\mathrm{n}}$ be a set, and $f : A \to \mathbb{R}$ be a real-valued function. The collection of points in A where f has a constant value is called a ***level set*** of f. The collection of all points $\big(x, f(x)\big)$ is called the ***graph*** of f.

Remark 2.6. A level surface is conventionally called a level curve when n = 2.

2.2 Limits and Continuity

Definition 2.7. Let $A \subseteq \mathbb{R}^{\mathrm{n}}$ be a set, and $f : A \to \mathbb{R}^{\mathrm{m}}$ be a vector-valued function. For a given $x_0 \in A'$, we say that $b \in \mathbb{R}^{\mathrm{m}}$ is the limit of f at x_0, written

$$\lim_{x \to x_0} f(x) = b \qquad \text{or} \qquad f(x) \to b \text{ as } x \to x_0 \,,$$

if for each $\varepsilon > 0$, there exists $\delta = \delta(x_0, \varepsilon) > 0$ such that

$$\|f(x) - b\|_{\mathbb{R}^{\mathrm{m}}} < \varepsilon \text{ whenever } 0 < \|x - x_0\|_{\mathbb{R}^{\mathrm{n}}} < \delta \text{ and } x \in A \,.$$

By the definition above, it is easy to see the following

Proposition 2.8. *Let* $A \subseteq \mathbb{R}^{\mathrm{n}}$ *be a set, and* $f, g : A \to \mathbb{R}^{\mathrm{m}}$ *be a vector-valued functions. Suppose that* $x_0 \in A'$, $f(x) = g(x)$ *for all* $x \in A \backslash \{x_0\}$, *and* $\lim\limits_{x \to x_0} f(x)$ *exists. Then* $\lim\limits_{x \to x_0} g(x)$ *exists and*

$$\lim_{x \to x_0} g(x) = \lim_{x \to x_0} f(x) \,.$$

The following proposition is standard, and we omit the proof.

Proposition 2.9. *Let* $A \subseteq \mathbb{R}^{\mathrm{n}}$ *be a set, and* $f, g : A \to \mathbb{R}^{\mathrm{m}}$ *be vector-valued functions,* $h : A \to \mathbb{R}$ *be a real-valued function. Suppose that* $x_0 \in A'$, *and* $\lim\limits_{x \to x_0} f(x) = a$, $\lim\limits_{x \to x_0} g(x) = b$, $\lim\limits_{x \to x_0} h(x) = c$. *Then*

$$\lim_{x \to x_0} (f + g)(x) = a + b \,, \qquad \lim_{x \to x_0} (f - g)(x) = a - b \,,$$
$$\lim_{x \to x_0} (hf)(x) = ca \,, \qquad \lim_{x \to x_0} (f \cdot g)(x) = a \cdot b \,,$$
$$\lim_{x \to x_0} \Big(\frac{f}{h}\Big) = \frac{a}{c} \quad \text{if } c \neq 0 \,.$$

Example 2.10. By Proposition 2.9,

$$\lim_{(x,y)\to(0,1)} \frac{x - xy + 3}{x^2y + 5xy - y^3} = \frac{0 - (0)(1) + 3}{(0)^2(1) + 5(0)(1) - (1)^3} = -3\,.$$

Example 2.11. Let $f : (0,\infty) \times (0,\infty) \to \mathbb{R}$ be given by $f(x,y) = \dfrac{x^2 - xy}{\sqrt{x} - \sqrt{y}}$. We cannot apply Proposition 2.9 to compute the limit $\lim\limits_{(x,y)\to(0,0)} f(x,y)$, if the limit exists, since $\lim\limits_{(x,y)\to(0,0)} (\sqrt{x} - \sqrt{y}) = 0$. Nevertheless, if $(x,y) \neq (0,0)$,

$$f(x,y) = \frac{x^2 - xy}{\sqrt{x} - \sqrt{y}} = \frac{x(x-y)(\sqrt{x} + \sqrt{y})}{(\sqrt{x} - \sqrt{y})(\sqrt{x} + \sqrt{y})} = x(\sqrt{x} + \sqrt{y})\,;$$

thus Proposition 2.8 and 2.9 imply that

$$\lim_{(x,y)\to(0,0)} f(x,y) = \lim_{(x,y)\to(0,0)} x(\sqrt{x} + \sqrt{y}) = 0\,.$$

Definition 2.12. Let $A \subseteq \mathbb{R}^n$ be a set, and $f : A \to \mathbb{R}^m$ be a vector-valued function. The function f is said to be continuous at $x_0 \in A \cap A'$ if $\lim\limits_{x\to x_0} f(x) = f(x_0)$. In other words, f is continuous at x_0 if

$$\forall\, \varepsilon > 0, \exists\, \delta = \delta(x_0,\varepsilon) > 0 \ni \|f(x) - f(x_0)\|_{\mathbb{R}^m} < \varepsilon \text{ whenever } \|x - x_0\|_{\mathbb{R}^n} < \delta \text{ and } x \in A\,.$$

If f is continuous at each point of $B \subseteq A \cap A'$, then f is said to be continuous on B.

Remark 2.13. 1. The notation $\delta = \delta(x_0,\varepsilon)$ means that the number δ could depend on x_0 and ε.

2. Another way of interpreting the continuity of f at x_0 is as follows: $f : A \to \mathbb{R}^m$ is continuous at $x_0 \in \mathcal{U}$ if

$$\forall\, \varepsilon > 0, \exists\, \delta = \delta(x_0,\varepsilon) > 0 \ni f\big(B(x_0,\delta) \cap A\big) \subseteq B(f(x_0),\varepsilon)\,.$$

3. If $A = \mathcal{U}$ is an open set, we can assume that δ is chosen small enough so that $B(x_0,\delta) \subseteq \mathcal{U}$ in both Definition 2.7 and 2.12. In other words, $\lim\limits_{x\to x_0} f(x) = b$ if

$$\forall\, \varepsilon > 0, \exists\, \delta = \delta(x_0,\varepsilon) > 0 \ni \|f(x) - b\|_{\mathbb{R}^m} < \varepsilon \text{ whenever } 0 < \|x - x_0\|_{\mathbb{R}^n} < \delta\,,$$

and $f : \mathcal{U} \to \mathbb{R}^m$ is continuous at $x_0 \in \mathcal{U}$ if

$$\forall\, \varepsilon > 0, \exists\, \delta = \delta(x_0,\varepsilon) > 0 \ni \|f(x) - f(x_0)\|_{\mathbb{R}^m} < \varepsilon \text{ whenever } \|x - x_0\|_{\mathbb{R}^n} < \delta\,.$$

4. If $A \subseteq \mathbb{R}^n$ is closed and bounded, and $f : A \to \mathbb{R}^m$ is continuous, then for each $\varepsilon > 0$ we can choose δ depending only on ε such that

$$\|f(x) - f(y)\|_{\mathbb{R}^m} < \varepsilon \text{ whenever } \|x - y\|_{\mathbb{R}^n} < \delta \text{ and } x, y \in A\,.$$

The property (that δ can be chosen independent of the point x_0) is called ***uniform continuity***.

Theorem 2.14. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}^m$ be a vector-valued function. Then the following assertions are equivalent:*

1. *f is continuous on $\mathcal{U}$.*

2. *For each open set $\mathcal{V} \subseteq \mathbb{R}^m$, $f^{-1}(\mathcal{V}) \subseteq \mathcal{U}$ is open, where $f^{-1}(\mathcal{V})$ is the pre-image of $\mathcal{V}$ under f defined by*
$$f^{-1}(\mathcal{V}) \equiv \big\{x \in \mathcal{U} \,\big|\, f(x) \in \mathcal{V}\big\}\,.$$

Proof. Before proceeding, we recall that $B \subseteq f^{-1}(f(B))$ for all $B \subseteq \mathcal{U}$ and $f(f^{-1}(B)) \subseteq B$ for all $B \subseteq \mathbb{R}^m$.

"1 $\Rightarrow$ 2" Let $a \in f^{-1}(\mathcal{V})$. Then $f(a) \in \mathcal{V}$. Since $\mathcal{V}$ is open in $\mathbb{R}^m$, $\exists\, \varepsilon_{f(a)} > 0$ such that $B(f(a), \varepsilon_{f(a)}) \subseteq \mathcal{V}$. By continuity of f (and Remark 2.13), there exists $\delta_a > 0$ such that

$$f\big(B(a, \delta_a)\big) \subseteq B\big(f(a), \varepsilon_{f(a)}\big)\,.$$

Therefore, for each $a \in f^{-1}(\mathcal{V})$, $\exists\, \delta_a > 0$ such that

$$B(a, \delta_a) \subseteq f^{-1}\big(f\big(B(a, \delta_a)\big)\big) \subseteq f^{-1}\big(B\big(f(a), \varepsilon_{f(a)}\big)\big) \subseteq f^{-1}(\mathcal{V})\,.$$

Therefore, $f^{-1}(\mathcal{V})$ is open.

"2 $\Rightarrow$ 1" Let $a \in \mathcal{U}$ and $\varepsilon > 0$ be given. Define $\mathcal{V} = B(f(a), \varepsilon)$, then $\mathcal{V}$ is open. Since $a \in f^{-1}(\mathcal{V})$ and $f^{-1}(\mathcal{V})$ is open by assumption, there exists $\delta > 0$ such that $B(a, \delta) \subseteq f^{-1}(\mathcal{V})$. Therefore,

$$f\big(B(a, \delta)\big) \subseteq f(f^{-1}(\mathcal{V})) \subseteq \mathcal{V} = B(f(a), \varepsilon)$$

which (with the help of Remark 2.13) implies that f is continuous at a. □

2.3 Definition of Derivatives and the Matrix Representation of Derivatives

Definition 2.15. Let $\mathcal{U} \subseteq \mathbb{R}^{\rm m}$ be an open set. A function $f : \mathcal{U} \to \mathbb{R}^{\rm m}$ is said to be ***differentiable*** at $x_0 \in A$ if there is a linear transformation from $\mathbb{R}^{\rm n}$ to $\mathbb{R}^{\rm m}$, denoted by $(Df)(x_0)$ and called the ***derivative*** of f at x_0, such that

$$\lim_{x \to x_0} \frac{\big\|f(x) - f(x_0) - (Df)(x_0)(x - x_0)\big\|_{\mathbb{R}^{\rm m}}}{\|x - x_0\|_{\mathbb{R}^{\rm n}}} = 0\,,$$

where $(Df)(x_0)(x - x_0)$ denotes the value of the linear transformation $(Df)(x_0)$ applied to the vector $x - x_0$. In other words, f is differentiable at $x_0 \in \mathcal{U}$ if there exists $L \in \mathscr{B}(\mathbb{R}^{\rm n}, \mathbb{R}^{\rm m})$ such that

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 \ni \|f(x) - f(x_0) - L(x - x_0)\|_{\mathbb{R}^{\rm m}} \leqslant \varepsilon \|x - x_0\|_{\mathbb{R}^{\rm n}} \text{ whenever } \|x - x_0\|_{\mathbb{R}^{\rm n}} < \delta\,.$$

If f is differentiable at each point of $\mathcal{U}$, we say that f is differentiable on $\mathcal{U}$.

Example 2.16. Let $L : \mathbb{R}^{\rm n} \to \mathbb{R}^{\rm m}$ be a linear transformation; that is, there is a matrix $[L]_{m\times n}$ such that $L(x) = [L]_{m\times n}[x]_n$ for all $x \in \mathbb{R}^{\rm n}$. Then L is differentiable. In fact, $(DL)(x_0) = L$ for all $x_0 \in X$ since

$$\lim_{x \to x_0} \frac{\|Lx - Lx_0 - L(x - x_0)\|_{\mathbb{R}^{\rm m}}}{\|x - x_0\|_{\mathbb{R}^{\rm n}}} = 0\,.$$

Example 2.17. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = x^2 + 2y$. Define $L_{(a,b)}(x, y) = 2ax + 2y$. Then $L_{(a,b)}$ is a linear transformation (from $\mathbb{R}^2$ to $\mathbb{R}$) and

$$\begin{aligned} &\frac{\big|x^2 + 2y - a^2 - 2b - L_{(a,b)}(x - a, y - b)\big|}{\sqrt{(x - a)^2 + (y - b)^2}} \\ &\qquad = \frac{\big|x^2 + 2y - a^2 - 2b - 2a(x - a) - 2(y - b)\big|}{\sqrt{(x - a)^2 + (y - b)^2}} \\ &\qquad = \frac{(x - a)^2}{\sqrt{(x - a)^2 + (y - b)^2}} \leqslant |x - a|\,; \end{aligned}$$

thus

$$\lim_{(x,y) \to (a,b)} \frac{\big|x^2 + 2y - a^2 - 2b - L_{(a,b)}(x - a, y - b)\big|}{\sqrt{(x - a)^2 + (y - b)^2}} = 0\,.$$

Therefore, f is differentiable at (a, b) and $(Df)(a, b) = L_{(a,b)}$.

Remark 2.18. Adopting the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$, a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ has a matrix representation $[L]_{m\times n}$ such that $L(x) = [L]_{m\times n}[x]_n$ for all $x \in \mathbb{R}^n$. In the following, we will always use the standard basis for $\mathbb{R}^n$ and $\mathbb{R}^m$ and use L and $L(x)$ to denote $[L]_{m\times m}$ and $[L]_{m\times n}[x]_n$, respectively, if L is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $x \in \mathbb{R}^n$.

Proposition 2.19. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be an open set, and $f : \mathcal{U} \to \mathbb{R}^m$ be differentiable at $x_0 \in \mathcal{U}$. Then $(Df)(x_0)$, the derivative of f at x_0, is uniquely determined by f.*

Proof. Suppose $L_1, L_2 \in \mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)$ are derivatives of f at x_0. Let $\varepsilon > 0$ be given and $e \in \mathbb{R}^n$ be a unit vector; that is, $\|e\|_{\mathbb{R}^n} = 1$. Since $\mathcal{U}$ is open, there exists $r > 0$ such that $B(x_0, r) \subseteq \mathcal{U}$. By Definition 2.15, there exists $0 < \delta < r$ such that

$$\frac{\|f(x) - f(x_0) - L_1(x - x_0)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} < \frac{\varepsilon}{2} \quad \text{and} \quad \frac{\|f(x) - f(x_0) - L_2(x - x_0)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} < \frac{\varepsilon}{2}$$

if $0 < \|x - x_0\|_{\mathbb{R}^n} < \delta$. Letting $x = x_0 + \lambda e$ with $0 < |\lambda| < \delta$, we have

$$\begin{aligned}
\|L_1 e - L_2 e\|_{\mathbb{R}^m} &= \frac{1}{|\lambda|}\|L_1(x - x_0) - L_2(x - x_0)\|_{\mathbb{R}^m} \\
&\leqslant \frac{1}{|\lambda|}\big(\big\|f(x) - f(x_0) - L_1(x - x_0)\big\|_{\mathbb{R}^m} + \big\|f(x) - f(x_0) - L_2(x - x_2)\big\|_{\mathbb{R}^m}\big) \\
&= \frac{\big\|f(x) - f(x_0) - L_1(x - x_0)\big\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} + \frac{\big\|f(x) - f(x_0) - L_2(x - x_0)\big\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $L_1 e = L_2 e$ for all unit vectors $e \in \mathbb{R}^n$ which guarantees that $L_1 = L_2$ (since if $x \neq 0$, $L_1 x = \|x\|_{\mathbb{R}^n} L_1\big(\frac{x}{\|x\|_{\mathbb{R}^n}}\big) = \|x\|_{\mathbb{R}^n} L_2\big(\frac{x}{\|x\|_{\mathbb{R}^n}}\big) = L_2 x$). ▫

Example 2.20. $(Df)(x_0)$ may not be unique if the domain of f is not open. For example, let $A = \big\{(x, y) \,\big|\, 0 \leqslant x \leqslant 1, y = 0\big\}$ be a subset of $\mathbb{R}^2$, and $f : A \to \mathbb{R}$ be given by $f(x, y) = 0$. Fix $x_0 = (a, 0) \in A$, then both of the linear maps

$$L_1(x, y) = 0 \quad \text{and} \quad L_2(x, y) = ay \qquad \forall\, (x, y) \in \mathbb{R}^2$$

satisfy Definition 2.15 since

$$\lim_{(x,0)\to(a,0)} \frac{\big|f(x, 0) - f(a, 0) - L_1(x - a, 0)\big|}{\big\|(x, 0) - (a, 0)\big\|_{\mathbb{R}^2}} = \lim_{(x,0)\to(a,0)} \frac{\big|f(x, 0) - f(a, 0) - L_2(x - a, 0)\big|}{\big\|(x, 0) - (a, 0)\big\|_{\mathbb{R}^2}} = 0\,.$$

Definition 2.21. Let $\{\mathrm{e}_k\}_{k=1}^{\mathrm{n}}$ be the standard basis of $\mathbb{R}^{\mathrm{n}}$, $\mathcal{U} \subseteq \mathbb{R}^{\mathrm{n}}$ be an open set, $a \in \mathcal{U}$ and $f : \mathcal{U} \to \mathbb{R}$ be a function. The partial derivative of f at a with respect to x_j, denoted by $\dfrac{\partial f}{\partial x_j}(a)$, is the limit

$$\lim_{h \to 0} \frac{f(a + h\mathrm{e}_j) - f(a)}{h}$$

if it exists. In other words, if $a = (a_1, \cdots, a_n)$, then

$$\frac{\partial f}{\partial x_j}(a) = \lim_{h \to 0} \frac{f(a_1, \cdots, a_{j-1}, a_j + h, a_{j+1}, \cdots, a_n) - f(a_1, \cdots, a_n)}{h}.$$

Theorem 2.22. *Suppose $\mathcal{U} \subseteq \mathbb{R}^{\mathrm{n}}$ is an open set and $f : \mathcal{U} \to \mathbb{R}^{\mathrm{m}}$ is differentiable at $a \in \mathcal{U}$. Then the partial derivatives $\dfrac{\partial f_i}{\partial x_j}(a)$ exists for all $i = 1, \cdots m$ and $j = 1, \cdots n$, and the matrix representation of the linear transformation $Df(a)$ (with respect to the standard basis of $\mathbb{R}^{\mathrm{n}}$ and $\mathbb{R}^{\mathrm{m}}$) is given by*

$$\big[Df(a)\big] = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(a) & \cdots & \dfrac{\partial f_1}{\partial x_{\mathrm{n}}}(a) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_{\mathrm{m}}}{\partial x_1}(a) & \cdots & \dfrac{\partial f_{\mathrm{m}}}{\partial x_{\mathrm{n}}}(a) \end{bmatrix} \quad \textit{or} \quad \big[Df(a)\big]_{ij} = \frac{\partial f_i}{\partial x_j}(a).$$

Proof. Since $\mathcal{U}$ is open and $a \in \mathcal{U}$, there exists $r > 0$ such that $B(a, r) \subseteq \mathcal{U}$. By the differentiability of f at a, there is $L \in \mathscr{B}(\mathbb{R}^{\mathrm{n}}, \mathbb{R}^{\mathrm{m}})$ such that for any given $\varepsilon > 0$, there exists $0 < \delta < r$ such that

$$\|f(x) - f(a) - L(x - a)\|_{\mathbb{R}^{\mathrm{m}}} \leqslant \varepsilon \|x - a\|_{\mathbb{R}^{\mathrm{n}}} \text{ whenever } x \in B(a, \delta).$$

In particular, for each $i = 1, \cdots, m$,

$$\Big|\frac{f_i(a + h\mathrm{e}_j) - f_i(a)}{h} - (L\mathrm{e}_j)_i\Big| \leqslant \Big\|\frac{f(a + h\mathrm{e}_j) - f(a)}{h} - L\mathrm{e}_j\Big\|_{\mathbb{R}^{\mathrm{m}}} \leqslant \varepsilon \quad \forall\, 0 < |h| < \delta, h \in \mathbb{R},$$

where $(L\mathrm{e}_j)_i$ denotes the i-th component of $L\mathrm{e}_j$ in the standard basis. As a consequence, for each $i = 1, \cdots, \mathrm{m}$,

$$\lim_{h \to 0} \frac{f_i(a + h\mathrm{e}_j) - f_i(a)}{h} = (L\mathrm{e}_j)_i \text{ exists}$$

and by definition, we must have $(L\mathrm{e}_j)_i = \dfrac{\partial f_i}{\partial x_j}(a)$. Therefore, $L_{ij} = \dfrac{\partial f_i}{\partial x_j}(a)$. □

Definition 2.23. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be an open set, and $f : \mathcal{U} \to \mathbb{R}^m$. The matrix

$$(Jf)(x) \equiv \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}(x) \equiv \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(x) & \cdots & \dfrac{\partial f_1}{\partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(x) & \cdots & \dfrac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

is called the ***Jacobian matrix*** of f at x (if each entry exists).

Remark 2.24. A function f might not be differential even if the Jacobian matrix Jf exists; however, if f is differentiable at x_0, then $(Df)(x)$ can be represented by $(Jf)(x)$; that is, $[(Df)(x)] = (Jf)(x)$.

Example 2.25. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(x_1, x_2) = (x_1^2, x_1^3 x_2, x_1^4 x_2^2)$. Suppose that f is differentiable at $x = (x_1, x_2)$, then

$$\big[(Df)(x)\big] = \begin{bmatrix} 2x_1 & 0 \\ 3x_1^2 x_2 & x_1^3 \\ 4x_1^3 x_2^2 & 2x_1^4 x_2 \end{bmatrix}.$$

Remark 2.26. For each $x \in A$, $Df(x)$ is a linear transformation, but Df in general is not linear in x.

Example 2.27. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0)\,, \\ 0 & \text{if } (x,y) = (0,0)\,. \end{cases}$$

Then $\dfrac{\partial f}{\partial x}(0,0) = \dfrac{\partial f}{\partial y}(0,0) = 0$; thus if f is differentiable at $(0,0)$, then $(Df)(0,0) = \begin{bmatrix} 0 & 0 \end{bmatrix}$. However,

$$\left| f(x,y) - f(0,0) - \begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right| = \frac{|xy|}{x^2+y^2} = \frac{|xy|}{(x^2+y^2)^{\frac{3}{2}}} \sqrt{x^2+y^2}\,;$$

thus f is not differentiable at $(0,0)$ since $\dfrac{|xy|}{(x^2+y^2)^{\frac{3}{2}}}$ cannot be arbitrarily small even if x^2+y^2 is small.

Example 2.28. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} x & \text{if } y = 0\,, \\ y & \text{if } x = 0\,, \\ 1 & \text{otherwise}\,. \end{cases}$$

Then $\dfrac{\partial f}{\partial x}(0,0) = \lim\limits_{h\to 0}\dfrac{f(h,0)-f(0,0)}{h} = \lim\limits_{h\to 0}\dfrac{h}{h} = 1$. Similarly, $\dfrac{\partial f}{\partial y}(0,0) = 1$; thus if f is differentiable at $(0,0)$, then $(Df)(0,0) = \begin{bmatrix}1 & 1\end{bmatrix}$. However,

$$\left|f(x,y) - f(0,0) - \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}\right| = \big|f(x,y)-(x+y)\big|;$$

thus if $xy \neq 0$,

$$\big|f(x,y)-(x+y)\big| = |1-x-y| \nrightarrow 0 \text{ as } (x,y)\to(0,0), xy\neq 0.$$

Therefore, f is not differentiable at $(0,0)$.

2.4 Conditions for Differentiability

Proposition 2.29. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $a \in \mathcal{U}$, and $f = (f_1,\cdots,f_m) : \mathcal{U} \to \mathbb{R}^m$. Then f is differentiable at a if and only if f_i is differentiable at a for all $i = 1,\cdots,m$. In other words, for vector-valued functions defined on an open subset of $\mathbb{R}^n$,*

Componentwise differentiable $\Leftrightarrow$ Differentiable.

Proof. "$\Rightarrow$" Let $(Df)(a)$ be the Jacobian matrix of f at a. Then

$$\forall\,\varepsilon>0, \exists\,\delta>0 \ni \big\|f(x)-f(a)-(Df)(a)(x-a)\big\|_{\mathbb{R}^m} \leqslant \varepsilon\|x-a\|_{\mathbb{R}^n} \text{ if } \|x-a\|_{\mathbb{R}^n}<\delta\,.$$

Let $\{\mathrm{e}_j\}_{j=1}^m$ be the standard basis of $\mathbb{R}^m$, and $L_i \in \mathscr{L}(\mathbb{R}^n,\mathbb{R})$ be given by $L_i(h) = \mathrm{e}_i^{\mathrm{T}}[(Df)(a)]h$. Then $L_i \in \mathscr{B}(\mathbb{R}^n,\mathbb{R})$ by Remark 1.79, and if $\|x-a\|_{\mathbb{R}^n}<\delta$,

$$\begin{aligned}\big|f_i(x)-f_i(a)-L_i(x-a)\big| &= \big|\mathrm{e}_i\cdot\big(f(x)-f(a)-(Df)(a)(x-a)\big)\big| \\ &\leqslant \big\|f(x)-f(a)-(Df)(a)(x-a)\big\|_{\mathbb{R}^m} \leqslant \varepsilon\|x-a\|_{\mathbb{R}^n};\end{aligned}$$

thus f_i is differentiable at a with derivatives L_i.

"$\Leftarrow$" Suppose that $f_i : \mathcal{U}\to\mathbb{R}$ is differentiable at a for each $i=1,\cdots,m$. Then there exists $L_i \in \mathscr{B}(\mathbb{R}^n,\mathbb{R})$ such that

$$\forall\,\varepsilon>0, \exists\,\delta_i>0 \ni \big|f_i(x)-f_i(a)-L_i(x-a)\big| \leqslant \frac{\varepsilon}{m}\|x-a\|_{\mathbb{R}^n} \text{ if } \|x-a\|_{\mathbb{R}^n}<\delta_i\,.$$

Let $L \in \mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ be given by $Lx = (L_1x, L_2x,\cdots,L_mx) \in \mathbb{R}^m$ if $x\in\mathbb{R}^n$. Then $L\in\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)$ by Remark 1.79, and

$$\big\|f(x)-f(a)-L(x-a)\big\|_{\mathbb{R}^m} \leqslant \sum_{i=1}^m \big|f_i(x)-f_i(a)-L_i(x-a)\big| \leqslant \varepsilon\|x-a\|_{\mathbb{R}^n}$$

if $\|x-a\|_{\mathbb{R}^n} < \delta = \min\big\{\delta_1,\cdots,\delta_m\big\}$. $\square$

Theorem 2.30. *Let $\mathcal{U} \subseteq \mathbb{R}^{\mathrm{n}}$ be open, $a \in \mathcal{U}$, and $f : \mathcal{U} \to \mathbb{R}$. If*

1. *the Jacobian matrix of f exists in a neighborhood of a, and*

2. *at least $(\mathrm{n}-1)$ entries of the Jacobian matrix of f are continuous at a,*

then f is differentiable at a.

Proof. W.L.O.G. we can assume that $\dfrac{\partial f}{\partial x_1}, \dfrac{\partial f}{\partial x_2}, \cdots, \dfrac{\partial f}{\partial x_{\mathrm{n}-1}}$ are continuous at a. Let $\{\mathrm{e}_j\}_{j=1}^{\mathrm{n}}$ be the standard basis of $\mathbb{R}^{\mathrm{n}}$, and $\varepsilon > 0$ be given. Since $\dfrac{\partial f}{\partial x_i}$ is continuous at a for $i = 1, \cdots, n-1$,

$$\exists\, \delta_i > 0 \ni \Big|\frac{\partial f}{\partial x_i}(x) - \frac{\partial f}{\partial x_i}(a)\Big| < \frac{\varepsilon}{\sqrt{n}} \text{ whenever } \|x-a\|_{\mathbb{R}^{\mathrm{n}}} < \delta_i\,.$$

On the other hand, by the definition of the partial derivatives,

$$\exists\, \delta_n > 0 \ni \Big|\frac{f(a+h\mathrm{e}_n)-f(a)}{h} - \frac{\partial f}{\partial x_n}(a)\Big| < \frac{\varepsilon}{\sqrt{n}} \text{ whenever } 0 < |h| < \delta_n\,.$$

Let $k = x - a$ and $\delta = \min\big\{\delta_1, \cdots, \delta_n\big\}$. Then

$$\begin{aligned}
&\Big|f(x) - f(a) - \Big[\frac{\partial f}{\partial x_1}(a)(x_1-a_1) + \cdots + \frac{\partial f}{\partial x_n}(a)(x_n-a_n)\Big]\Big| \\
&\quad = \Big|f(a+k) - f(a) - \frac{\partial f}{\partial x_1}(a)k_1 - \cdots - \frac{\partial f}{\partial x_n}(a)k_n\Big| \\
&\quad = \Big|f(a_1+k_1, \cdots, a_n+k_n) - f(a_1, \cdots, a_n) - \frac{\partial f}{\partial x_1}(a)k_1 - \cdots - \frac{\partial f}{\partial x_n}(a)k_n\Big| \\
&\quad \leqslant \Big|f(a_1+k_1, \cdots, a_n+k_n) - f(a_1, a_2+k_2, \cdots, a_n+k_n) - \frac{\partial f}{\partial x_1}(a)k_1\Big| \\
&\qquad + \Big|f(a_1, a_2+k_2, \cdots, a_n+k_n) - f(a_1, a_2, a_3+k_3, \cdots, a_n+k_n) - \frac{\partial f}{\partial x_2}(a)k_2\Big| \\
&\qquad + \cdots + \Big|f(a_1, \cdots, a_{n-1}, a_n+k_n) - f(a_1, \cdots, a_n) - \frac{\partial f}{\partial x_n}(a)k_n\Big|\,.
\end{aligned}$$

By the mean value theorem,

$$\begin{aligned}
&f(a_1, \cdots, a_{j-1}, a_j+k_j, \cdots, a_n+k_n) - f(a_1, \cdots, a_j, a_{j+1}+k_{j+1}, \cdots, a_n+k_n) \\
&\quad = k_j \frac{\partial f}{\partial x_j}(a_1, \cdots, a_{j-1}, a_j+\theta_j k_j, a_{j+1}+k_{j+1}, \cdots, a_n+k_n)
\end{aligned}$$

for some $0 < \theta_j < 1$; thus for $j = 1, \cdots, n-1$, if $\|x-a\|_{\mathbb{R}^{\mathrm{n}}} = \|k\|_{\mathbb{R}^{\mathrm{n}}} < \delta$,

$$\begin{aligned}
&\Big|f(a_1, \cdots, a_{j-1}, a_j+k_j, \cdots, a_n+k_n) - f(a_1, \cdots, a_j, a_{j+1}+k_{j+1}, \cdots, a_n+k_n) - \frac{\partial f}{\partial x_j}(a)k_j\Big| \\
&\quad = \Big|\frac{\partial f}{\partial x_j}(a_1, \cdots, a_{j-1}, a_j+\theta_j k_j, a_{j+1}+k_{j+1}, \cdots, a_n+k_n) - \frac{\partial f}{\partial x_j}(a)\Big||k_j| \leqslant \frac{\varepsilon}{\sqrt{n}}|k_j|\,.
\end{aligned}$$

Moreover, if $\|x-a\|_{\mathbb{R}^n} < \delta$, then $|k_n| \leqslant \|k\|_{\mathbb{R}^n} = \|x-a\|_{\mathbb{R}^n} < \delta \leqslant \delta_n$; thus

$$\Big|f(a_1,\cdots,a_{n-1},a_n+k_n) - f(a_1,\cdots,a_n) - \frac{\partial f}{\partial x_n}(a)k_n\Big| \leqslant \frac{\varepsilon}{\sqrt{n}}|k_n|\,.$$

As a consequence, if $\|x-a\|_{\mathbb{R}^n} < \delta$, by Cauchy's inequality,

$$\begin{aligned}&\Big|f(x)-f(a)-\Big[\frac{\partial f}{\partial x_1}(a)(x_1-a_1)+\cdots+\frac{\partial f}{\partial x_n}(a)(x_n-a_n)\Big]\Big|\\ &\qquad\leqslant \frac{\varepsilon}{\sqrt{n}}\sum_{j=1}^{n}|k_j| \leqslant \varepsilon\|k\|_{\mathbb{R}^n} = \varepsilon\|x-a\|_{\mathbb{R}^n}\end{aligned}$$

which implies that f is differentiable at a. □

Remark 2.31. When two or more components of the Jacobian matrix $\Big[\frac{\partial f}{\partial x_1}\ \cdots\ \frac{\partial f}{\partial x_n}\Big]$ of a scalar function f are discontinuous at a point $x_0 \in \mathcal{U}$, in general f is not differentiable at x_0. For example, both components of the Jacobian matrix of the functions given in Example 2.27, 2.28, 2.44 are discontinuous at $(0,0)$, and these functions are not differentiable at $(0,0)$.

Example 2.32. Let $\mathcal{U} = \mathbb{R}^2 \backslash \big\{(x,0)\in\mathbb{R}^2 \,\big|\, x \geqslant 0\big\}$, and $f:\mathcal{U}\to\mathbb{R}$ be given by

$$f(x,y) = \arg(x+iy) = \begin{cases} \cos^{-1}\dfrac{x}{\sqrt{x^2+y^2}} & \text{if } y>0\,,\\ \pi & \text{if } y=0\,,\\ 2\pi-\cos^{-1}\dfrac{x}{\sqrt{x^2+y^2}} & \text{if } y<0\,.\end{cases}$$

Then

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} -\dfrac{y}{x^2+y^2} & \text{if } y\neq 0\,,\\ 0 & \text{if } y=0\,,\end{cases} \quad\text{and}\quad \frac{\partial f}{\partial y}(x,y) = \begin{cases} \dfrac{x}{x^2+y^2} & \text{if } y\neq 0\,,\\ \dfrac{1}{x} & \text{if } y=0\,.\end{cases}$$

Since $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ are both continuous on $\mathcal{U}$, f is differentiable on $\mathcal{U}$.

Definition 2.33. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f:\mathcal{U}\to\mathbb{R}^m$ be differentiable on $\mathcal{U}$. f is said to be ***continuously differentiable*** on $\mathcal{U}$ if the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist and are continuous on $\mathcal{U}$ for $i=1,\cdots,m$ and $j=1,\cdots,n$. The collection of all continuously differentiable functions from $\mathcal{U}$ to $\mathbb{R}^m$ is denoted by $\mathscr{C}^1(\mathcal{U};\mathbb{R}^m)$. The collection of all bounded differentiable functions from $\mathcal{U}$ to $\mathbb{R}^m$ whose partial derivatives are continuous and bounded is denoted by $\mathscr{C}^1_b(\mathcal{U};\mathbb{R}^m)$.

Example 2.34. If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at x_0, must f' be continuous at x_0? In other words, is it always true that $\lim\limits_{x \to x_0} f'(x) = f'(x_0)$?

Answer: No! For example, take

$$f(x) = \begin{cases} x^2 \sin \dfrac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

1° Show $f(x)$ is differentiable at $x = 0$:

$$f'(0) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0} \frac{h^2 \sin \frac{1}{h}}{h} = \lim_{h \to 0} h \sin \frac{1}{h} = 0\,.$$

2° We compute the derivative of f and find that

$$f'(x) = \begin{cases} 2x \sin \dfrac{1}{x} - \cos \dfrac{1}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

However, $\lim\limits_{x \to 0} f'(x)$ does not exist.

Definition 2.35. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$ be a function. If the partial derivative $\dfrac{\partial f}{\partial x_j}$ exists in $\mathcal{U}$ and has partial derivatives (at every point in $\mathcal{U}$) with respect to x_i, then the second-order partial derivatives $\dfrac{\partial}{\partial x_i}\Big(\dfrac{\partial f}{\partial x_j}\Big)$ is denoted by $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$.

In general, if the k-th order partial derivatives $\dfrac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}$ exists in $\mathcal{U}$ and has partial derivatives (at every point in $\mathcal{U}$) with respect to $x_{i_{k+1}}$, then the $(k+1)$-th order partial derivatives $\dfrac{\partial}{\partial x_{i_{k+1}}}\Big(\dfrac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}\Big)$ is denoted by $\dfrac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \cdots \partial x_{i_1}}$; that is,

$$\frac{\partial^{k+1} f}{\partial x_{i_{k+1}} \partial x_{i_k} \cdots \partial x_{i_1}} = \frac{\partial}{\partial x_{i_{k+1}}}\Big(\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}\Big)\,.$$

Theorem 2.36. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $a \in \mathcal{U}$, and $f : \mathcal{U} \to \mathbb{R}$ be a real-valued function. Suppose that for some $1 \leqslant i, j \leqslant n$, $\dfrac{\partial f}{\partial x_i}$, $\dfrac{\partial f}{\partial x_j}$, $\dfrac{\partial^2 f}{\partial x_j \partial x_i}$ and $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$ exist in a neighborhood of a and are continuous at a. Then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)\,.$$

Proof. W.L.O.G., we assume that f is a function of two variables; that is, n = 2. For fixed $h, k \in \mathbb{R}$, define $\varphi(x,y) = f(x, y+k) - f(x,y)$ and $\psi(x,y) = f(x+h, y) - f(x,y)$. Then

$$\begin{aligned}\varphi(a+h,b) - \varphi(a,b) &= f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)\\ &= \psi(a,b+k) - \psi(a,b)\,.\end{aligned}$$

By the mean value theorem (Theorem A.9), for $h, k \neq 0$ and sufficiently small,

$$\begin{aligned}\varphi(a+h,b) - \varphi(a,b) = \varphi_x(a+\theta_1 h, b)h &= \big[f_x(a+\theta_1 h, b+k) - f_x(a+\theta_1 h, b)\big]h\\ &= (f_x)_y(a+\theta_1 h, b+\theta_2 k)hk\end{aligned}$$

for some $\theta_1, \theta_2 \in (0,1)$, and similarly, for some $\theta_3, \theta_4 \in (0,1)$,

$$\psi(a,b+k) - \psi(a,b) = (f_y)_x(a+\theta_3 h, b+\theta_4 k)hk\,.$$

Therefore, for $h, k \neq 0$ and sufficiently small, there exist $\theta_1, \theta_2, \theta_3, \theta_4 \in (0,1)$ such that

$$(f_x)_y(a+\theta_1 h, b+\theta_2 k) = (f_y)_x(a+\theta_3 h, b+\theta_4 k)\,. \tag{2.1}$$

Let $\varepsilon > 0$ be given. Since $(f_x)_y$ and $(f_y)_x$ are continuous at (a,b), there exist $\delta_1, \delta_2 > 0$ such that

$$\begin{aligned}\big|(f_x)_y(x,y) - (f_x)_y(a,b)\big| &< \frac{\varepsilon}{2} \quad \text{if} \quad \sqrt{(x-a)^2 + (y-b)^2} < \delta_1\,,\\ \big|(f_x)_y(x,y) - (f_x)_y(a,b)\big| &< \frac{\varepsilon}{2} \quad \text{if} \quad \sqrt{(x-a)^2 + (y-b)^2} < \delta_2\,.\end{aligned}$$

In particular, if $\delta = \min\{\delta_1, \delta_2\}$ and $h, k \neq 0$ satisfying $\sqrt{h^2+k^2} < \delta$,

$$\big|(f_x)_y(a+\theta_1 h, b+\theta_2 k) - (f_x)_y(a,b)\big| + \big|(f_x)_y(a+\theta_3 h, b+\theta_4 k) - (f_x)_y(a,b)\big| < \varepsilon\,,$$

where $\theta_1, \theta_2, \theta_3, \theta_4 \in (0,1)$ are chosen to validate (2.1). As a consequence,

$$\begin{aligned}&\big|(f_x)_y(a,b) - (f_y)_x(a,b)\big|\\ &\quad = \big|(f_x)_y(a,b) - (f_x)_y(a+\theta_1 h, b+\theta_2 k) + (f_x)_y(a+\theta_3 h, b+\theta_4 k) - (f_x)_y(a,b)\big|\\ &\quad \leqslant \big|(f_x)_y(a+\theta_1 h, b+\theta_2 k) - (f_x)_y(a,b)\big| + \big|(f_x)_y(a+\theta_3 h, b+\theta_4 k) - (f_x)_y(a,b)\big| < \varepsilon\end{aligned}$$

which concludes the theorem (since $\varepsilon > 0$ is given arbitrarily). □

Example 2.37. Let $f:\mathbb{R}^2\to\mathbb{R}$ be defined by

$$f(x,y)=\begin{cases}\dfrac{xy(x^2-y^2)}{x^2+y^2} & \text{if } (x,y)\neq(0,0)\,,\\ 0 & \text{if } (x,y)=(0,0)\,.\end{cases}$$

Then

$$f_x(x,y)=\begin{cases}\dfrac{x^4y+4x^2y^3-y^5}{(x^2+y^2)^2} & \text{if } (x,y)\neq(0,0)\,,\\ 0 & \text{if } (x,y)=(0,0)\,,\end{cases}$$

and

$$f_y(x,y)=\begin{cases}\dfrac{x^5-4x^3y^2-xy^4}{(x^2+y^2)^2} & \text{if } (x,y)\neq(0,0)\,,\\ 0 & \text{if } (x,y)=(0,0)\,,\end{cases}$$

It is clear that f_x and f_y are continuous on $\mathbb{R}^2$; thus f is differentiable on $\mathbb{R}^2$. However,

$$f_{xy}(0,0)=\lim_{k\to0}\frac{f_x(0,k)-f_x(0,0)}{k}=-1\,,$$

while

$$f_{yx}(0,0)=\lim_{h\to0}\frac{f_y(h,0)-f_y(0,0)}{h}=1\,;$$

thus the Hessian matrix of f at the origin is not symmetric.

Definition 2.38. Let $\mathcal{U}\subseteq\mathbb{R}^n$ be open, and $f:\mathcal{U}\to\mathbb{R}^m$ be a vector-valued function. The function f is said to be of class $\mathscr{C}^2$ if $f\in\mathscr{C}^1(\mathcal{U};\mathbb{R}^m)$ and the second partial derivatives $\dfrac{\partial^2 f_i}{\partial x_j\partial x_k}$ exists and is continuous in $\mathcal{U}$ for all $1\leqslant i\leqslant m$ and $1\leqslant j,k\leqslant n$. The collection of all $\mathscr{C}^2$-functions $f:\mathcal{U}\to\mathbb{R}^m$ is denoted by $\mathscr{C}^2(\mathcal{U};\mathbb{R}^m)$.

In general, the function f is said to be of class $\mathscr{C}^k$ if $f\in\mathscr{C}^{k-1}(\mathcal{U};\mathbb{R}^m)$ and the k-th order partial derivatives $\dfrac{\partial^k f}{\partial x_{i_k}\partial x_{i_{k-1}}\cdots\partial x_{i_1}}$ exists and is continuous in $\mathcal{U}$ for all $1\leqslant i\leqslant m$ and $1\leqslant i_1,\cdots,i_k\leqslant n$. The collection of all $\mathscr{C}^k$-functions $f:\mathcal{U}\to\mathbb{R}^m$ is denoted by $\mathscr{C}^k(\mathcal{U};\mathbb{R}^m)$.

A function is said to be ***smooth*** or ***of class*** $\mathscr{C}^\infty$ if it is of class $\mathscr{C}^k$ for all positive integer k.

Corollary 2.39. *Let $\mathcal{U}\subseteq\mathbb{R}^n$ be open, and $f\in\mathscr{C}^2(\mathcal{U};\mathbb{R})$. Then*

$$\frac{\partial^2 f}{\partial x_i\partial x_j}(a)=\frac{\partial^2 f}{\partial x_j\partial x_i}(a)\qquad\forall\, a\in\mathcal{U}\ \text{and}\ 1\leqslant i,j\leqslant n\,.$$

2.5 Properties of Differentiable Functions

2.5.1 Continuity of Differentiable Functions

Theorem 2.40. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}^m$ be differentiable at $x_0 \in \mathcal{U}$. Then f is continuous at x_0.*

Proof. Since f is differentiable at x_0, there exists $L \in \mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)$ such that

$$\exists\, \delta_1 > 0 \ni \big\| f(x) - f(x_0) - L(x - x_0) \big\|_{\mathbb{R}^m} \leqslant \|x - x_0\|_{\mathbb{R}^n} \quad \forall\, x \in B(x_0, \delta_1)\,.$$

As a consequence,

$$\big\| f(x) - f(x_0) \big\|_{\mathbb{R}^m} \leqslant \big(\|L\| + 1\big) \|x - x_0\|_{\mathbb{R}^n} \quad \forall\, x \in B(x_0, \delta_1)\,. \tag{2.2}$$

For a given $\varepsilon > 0$, let $\delta = \min\Big\{\delta_1, \dfrac{\varepsilon}{2(\|L\| + 1)}\Big\}$. Then $\delta > 0$, and if $x \in B(x_0, \delta)$,

$$\big\| f(x) - f(x_0) \big\|_{\mathbb{R}^m} \leqslant \frac{\varepsilon}{2} < \varepsilon\,.$$ □

Remark 2.41. In fact, if f is differentiable at x_0, then f satisfies the "local Lipschitz property"; that is,

$$\exists\, M = M(x_0) > 0 \text{ and } \delta = \delta(x_0) > 0 \ni \text{if } \|x - x_0\|_X < \delta, \text{ then } \|f(x) - f(x_0)\|_Y \leqslant M\|x - x_0\|_X$$

since we can choose $M = \|L\| + 1$ and $\delta = \delta_1$ (see (2.2)).

Example 2.42. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given in Example 2.27. We have shown that f is not differentiable at $(0, 0)$. In fact, f is not even continuous at $(0, 0)$ since when approaching the origin along the straight line $x_2 = mx_1$,

$$\lim_{(x_1, mx_1) \to (0,0)} f(x_1, mx_1) = \lim_{x_1 \to 0} \frac{mx_1^2}{(m^2 + 1)x_1^2} = \frac{m^2}{m^2 + 1} \neq f(0, 0) \text{ if } m \neq 0\,.$$

Example 2.43. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given in Example 2.28. Then f is not continuous at $(0, 0)$; thus not differentiable at $(0, 0)$.

Example 2.44. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = \begin{cases} \dfrac{x^3}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0)\,, \\ 0 & \text{if } (x, y) = (0, 0)\,. \end{cases}$$

Then $f_x(0,0) = 1$ and $f_y(0,0) = 0$. However,

$$\frac{\left|f(x,y) - f(0,0) - \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}\right|}{\sqrt{x^2+y^2}} = \frac{|x|y^2}{(x^2+y^2)^{\frac{3}{2}}} \nrightarrow 0 \text{ as } (x,y) \to (0,0).$$

Therefore, f is not differentiable at $(0,0)$. On the other hand, f is continuous at $(0,0)$ since

$$\big|f(x,y) - f(0,0)\big| = \big|f(x,y)\big| \leqslant |x| \to 0 \text{ as } (x,y) \to (0,0).$$

2.5.2 The Product Rules

Proposition 2.45. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be an open set, and $f : \mathcal{U} \to \mathbb{R}^m$ and $g : \mathcal{U} \to \mathbb{R}$ be differentiable at $x_0 \in A$. Then $gf : A \to \mathbb{R}^m$ is differentiable at x_0, and*

$$D(gf)(x_0)(v) = g(x_0)(Df)(x_0)(v) + (Dg)(x_0)(v)f(x_0)\,. \tag{2.3}$$

Moreover, if $g(x_0) \neq 0$, then $\dfrac{f}{g} : A \to \mathbb{R}^m$ is also differentiable at x_0, and $D(\dfrac{f}{g})(x_0) : \mathbb{R}^n \to \mathbb{R}^m$ is given by

$$D\Big(\frac{f}{g}\Big)(x_0)(v) = \frac{g(x_0)\big((Df)(x_0)(v)\big) - (Dg)(x_0)(v)f(x_0)}{g^2(x_0)}\,. \tag{2.4}$$

Proof. We only prove (2.3), and (2.4) is left as an exercise.

Let A be the Jacobian matrix of gf at x_0; that is, the (i,j)-th entry of A is

$$\frac{\partial(gf_i)}{\partial x_j}(x_0) = g(x_0)\frac{\partial f_i}{\partial x_j}(x_0) + \frac{\partial g}{\partial x_j}(x_0)f_i(x_0)\,.$$

Then $Av = g(x_0)(Df)(x_0)(v) + (Dg)(x_0)(v)f(x_0)$; thus

$$\begin{aligned}(gf)(x) - (gf)(x_0) - A(x-x_0) &= g(x_0)\big(f(x) - f(x_0) - (Df)(x_0)(x-x_0)\big)\\ &\quad + \big(g(x) - g(x_0) - (Dg)(x_0)(x-x_0)\big)f(x)\\ &\quad + \big((Dg)(x_0)(x-x_0)\big)\big(f(x) - f(x_0)\big)\,.\end{aligned}$$

Since $(Dg)(x_0) \in \mathscr{B}(\mathbb{R}^n,\mathbb{R})$, $\|(Dg)(x_0)\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R})} < \infty$; thus using the inequality

$$\big|(Dg)(x_0)(x-x_0)\big| \leqslant \big\|(Dg)(x_0)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R})}\|x-x_0\|_{\mathbb{R}^n}$$

and the continuity of f at x_0 (due to Theorem 2.40), we find that

$$\lim_{x\to x_0}\left|\frac{\big|(Dg)(x_0)(x-x_0)\big|}{\|x-x_0\|_{\mathbb{R}^n}}\big\|f(x)-f(x_0)\big\|_{\mathbb{R}^m}\right| \leqslant \lim_{x\to x_0}\big\|(Dg)(x_0)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R})}\big\|f(x)-f(x_0)\big\|_{\mathbb{R}^m} = 0\,.$$

As a consequence,

$$\begin{aligned}
&\lim_{x\to x_0}\frac{\big\|(gf)(x)-(gf)(x_0)-A(x-x_0)\big\|_{\mathbb{R}^m}}{\|x-x_0\|_{\mathbb{R}^n}}\\
&\quad\leqslant \big|g(x_0)\big|\lim_{x\to x_0}\frac{\big\|f(x)-f(x_0)-(Df)(x_0)(x-x_0)\big\|_{\mathbb{R}^m}}{\|x-x_0\|_{\mathbb{R}^n}}\\
&\qquad+\lim_{x\to x_0}\Big[\frac{\big|g(x)-g(x_0)-(Dg)(x_0)(x-x_0)\big|}{\|x-x_0\|_{\mathbb{R}^n}}\|f(x)\|_{\mathbb{R}^m}\Big]\\
&\qquad+\lim_{x\to x_0}\Big[\frac{\big|(Dg)(x_0)(x-x_0)\big|}{\|x-x_0\|_{\mathbb{R}^n}}\big\|f(x)-f(x_0)\big\|_{\mathbb{R}^m}\Big]=0
\end{aligned}$$

which implies that gf is differentiable at x_0 with derivative $D(gf)(x_0)$ given by (2.3). □

- **The differentiation of the Jacobian**

Before going into the next section, we study the differentiation of a special determinant, the Jacobian.

Example 2.46. Suppose that $\psi : \Omega \subseteq \mathbb{R}^n \to \psi(\Omega) \subseteq \mathbb{R}^n$ is a given diffeomorphism (thus $\det(\nabla\psi) \neq 0$). Let $M = \nabla\psi$, and $J = \det(M)$. By Corollary 1.72, the adjoint matrix of M is JM^{-1}. Letting δ be a (first order) partial differential operator which satisfies $\delta(fg) = f\delta g + (\delta f)g$, by Theorem 1.73 we find that

$$\delta J = \operatorname{tr}(JM^{-1}\delta M) = \sum_{i,j=1}^{n} JA_i^j \delta\psi^i_{,j} \overset{\text{Einstein's summation convention}}{\equiv} JA_i^j\delta\psi^i_{,j}\,, \tag{2.5}$$

where $A_i^j = a_{ji}$ with $M^{-1} = [a_{ji}]_{n\times n}$, and $f_{,j} \equiv \dfrac{\partial f}{\partial x_j}$.

Remark 2.47. From now on we sometimes write the row index of a matrix as a super-script for the following reason: if $\psi : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a differentiable vector-valued function, then $\nabla\psi$ is usually expressed by

$$\nabla\psi = \begin{bmatrix} \dfrac{\partial\psi_1}{\partial x_1} & \dfrac{\partial\psi_1}{\partial x_2} & \cdots & \dfrac{\partial\psi_1}{\partial x_n} \\ \dfrac{\partial\psi_2}{\partial x_1} & \dfrac{\partial\psi_2}{\partial x_2} & \cdots & \dfrac{\partial\psi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial\psi_m}{\partial x_1} & \dfrac{\partial\psi_m}{\partial x_2} & \cdots & \dfrac{\partial\psi_m}{\partial x_n} \end{bmatrix};$$

thus the (i,j) element of $\nabla\psi$ is $\dfrac{\partial\psi_i}{\partial x_j}$, and the row index i appears "above" the column index j.

Theorem 2.48 (Piola's identity). *Let $\psi : \Omega \subseteq \mathbb{R}^{\mathrm{n}} \to \psi(\Omega) \subseteq \mathbb{R}^{\mathrm{n}}$ be a $\mathscr{C}^2$-diffeomorphism, and $[a_{ij}]_{\mathrm{n}\times\mathrm{n}}$ be the adjoint matrix of $\nabla\psi$. Then*

$$a_{ji},_j \overset{\substack{\text{Einstein's summation}\\ \text{convention}}}{\equiv} \sum_{j=1}^{\mathrm{n}} \frac{\partial}{\partial x_j} a_{ji} = 0. \tag{2.6}$$

In other words, each column of the adjoint matrix of the Jacobian matrix of ψ is divergence-free (see Definition 4.74).

Proof. Let $\mathrm{J} = \det(\nabla\psi)$ and $\mathrm{A} = (\nabla\psi)^{-1}$. Then $a_{ji} = \mathrm{J}\mathrm{A}^j_i$. Moreover, since $\mathrm{A}\nabla\psi = \mathrm{I}_n$, $\sum\limits_{r=1}^{\mathrm{n}} \mathrm{A}^j_r \psi^r_{,s} = \delta_{js}$; thus

$$0 = \Big[\sum_{r=1}^{\mathrm{n}} \mathrm{A}^j_r \psi^r_{,s}\Big]_{,k} = \sum_{r=1}^{\mathrm{n}} \big[\mathrm{A}^j_{r,k}\psi^r_{,s} + \mathrm{A}^j_r \psi^r_{,sk}\big]$$

which, after multiplying the equality above by A^s_i and then summing over s, implies that

$$\mathrm{A}^j_{i,k} = -\sum_{r,s=1}^{\mathrm{n}} \mathrm{A}^j_r \psi^r_{,sk} \mathrm{A}^s_i. \tag{2.7}$$

As a consequence, by Theorem 2.36 we conclude that

$$\sum_{j=1}^{\mathrm{n}} \frac{\partial}{\partial x_j}(\mathrm{J}\mathrm{A}^j_i) = \sum_{j=1}^{\mathrm{n}} \sum_{r,s=1}^{\mathrm{n}} \Big[\mathrm{J}\mathrm{A}^r_s \psi^s_{,rj} \mathrm{A}^j_i - \mathrm{J}\mathrm{A}^j_r \psi^r_{,sj} \mathrm{A}^s_i\Big] = 0\,. \qquad \square$$

2.5.3 The Chain Rule

Theorem 2.49. *Let $\mathcal{U} \subseteq \mathbb{R}^{\mathrm{n}}$ and $\mathcal{V} \subseteq \mathbb{R}^{\mathrm{m}}$ be open sets, $f : \mathcal{U} \to \mathbb{R}^{\mathrm{m}}$ and $g : \mathcal{V} \to \mathbb{R}^{\ell}$ be vector-valued functions, and $f(\mathcal{U}) \subseteq \mathcal{V}$. If f is differentiable at $x_0 \in \mathcal{U}$ and g is differentiable at $f(x_0)$, then the map $F = g \circ f$ defined by*

$$F(x) = g\big(f(x)\big) \qquad \forall\, x \in \mathcal{U}$$

is differentiable at x_0, and

$$(DF)(x_0)(h) = (Dg)\big(f(x_0)\big)\big((Df)(x_0)(h)\big)$$

or in component,

$$\big[(DF)(x_0)\big]_{ij} = \sum_{k=1}^{\mathrm{m}} \frac{\partial g_i}{\partial y_k}\big(f(x_0)\big)\frac{\partial f_k}{\partial x_j}(x_0)\,.$$

Proof. To simplify the notation, let $y_0 = f(x_0)$, $A = (Df)(x_0) \in \mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)$, and $B = (Dg)(y_0) \in \mathscr{B}(\mathbb{R}^m, \mathbb{R}^\ell)$. Let $\varepsilon > 0$ be given. By the differentiability of f and g at x_0 and y_0, there exists $\delta_1, \delta_2 > 0$ such that if $\|x - x_0\|_{\mathbb{R}^n} < \delta_1$ and $\|y - y_0\|_{\mathbb{R}^m} < \delta_2$, we have

$$\|f(x) - f(x_0) - A(x - x_0)\|_{\mathbb{R}^m} \leqslant \min\big\{1, \frac{\varepsilon}{2(\|B\|+1)}\big\}\|x - x_0\|_{\mathbb{R}^n},$$
$$\|g(y) - g(y_0) - B(y - y_0)\|_{\mathbb{R}^\ell} \leqslant \frac{\varepsilon}{2(\|A\|+1)}\|y - y_0\|_{\mathbb{R}^m}.$$

Define

$$u(h) = f(x_0 + h) - f(x_0) - Ah \qquad \forall\, \|h\|_{\mathbb{R}^n} < \delta_1,$$
$$v(k) = g(y_0 + k) - g(y_0) - Bk \qquad \forall\, \|k\|_{\mathbb{R}^m} < \delta_2.$$

Then if $\|h\|_{\mathbb{R}^n} < \delta_1$ and $\|k\|_{\mathbb{R}^m} < \delta_2$,

$$\|u(h)\|_{\mathbb{R}^m} \leqslant \|h\|_{\mathbb{R}^n}, \quad \|u(h)\|_{\mathbb{R}^m} \leqslant \frac{\varepsilon}{2(\|B\|+1)}\|h\|_{\mathbb{R}^n} \quad \text{and} \quad \|v(k)\|_{\mathbb{R}^\ell} \leqslant \frac{\varepsilon}{2(\|A\|+1)}\|k\|_{\mathbb{R}^m}.$$

Let $k = f(x_0 + h) - f(x_0) = Ah + u(h)$. Then $\lim\limits_{h\to 0} k = 0$; thus there exists $\delta_3 > 0$ such that

$$\|k\|_{\mathbb{R}^m} < \delta_2 \quad \text{whenever} \quad \|h\|_{\mathbb{R}^n} < \delta_3.$$

Since

$$F(x_0 + h) - F(x_0) = g(y_0 + k) - g(y_0) = Bk + v(k) = B\big(Ah + u(h)\big) + v(k)$$
$$= BAh + Bu(h) + v(k),$$

we conclude that if $\|h\|_{\mathbb{R}^n} < \delta = \min\{\delta_1, \delta_3\}$,

$$\|F(x_0 + h) - F(x_0) - BAh\|_{\mathbb{R}^\ell} \leqslant \|Bu(h)\|_{\mathbb{R}^\ell} + \|v(k)\|_{\mathbb{R}^\ell} \leqslant \|B\|\|u(h)\|_{\mathbb{R}^m} + \frac{\varepsilon}{2(\|A\|+1)}\|k\|_{\mathbb{R}^m}$$
$$\leqslant \frac{\varepsilon}{2}\|h\|_{\mathbb{R}^n} + \frac{\varepsilon}{2(\|A\|+1)}\big(\|A\|\|h\|_{\mathbb{R}^n} + \|u(h)\|_{\mathbb{R}^m}\big) \leqslant \frac{\varepsilon}{2}\|h\|_{\mathbb{R}^n} + \frac{\varepsilon}{2}\|h\|_{\mathbb{R}^n} = \varepsilon\|h\|_{\mathbb{R}^n}$$

which implies that F is differentiable at x_0 and $\big[(DF)(x_0)\big] = BA$. □

Example 2.50. Consider the polar coordinate $x = r\cos\theta$, $y = r\sin\theta$. Then every function $f : \mathbb{R}^2 \to \mathbb{R}$ is associated with a function $F : [0,\infty) \times [0, 2\pi) \to \mathbb{R}$ satisfying

$$F(r, \theta) = f(r\cos\theta, r\sin\theta).$$

Suppose that f is differentiable. Then F is differentiable, and the chain rule implies that

$$\begin{bmatrix} \dfrac{\partial F}{\partial r} & \dfrac{\partial F}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}.$$

Therefore, we arrive at the following form of chain rule

$$\frac{\partial}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial}{\partial x} + \frac{\partial y}{\partial r}\frac{\partial}{\partial y} \qquad \text{and} \qquad \frac{\partial}{\partial \theta} = \frac{\partial x}{\partial \theta}\frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta}\frac{\partial}{\partial y}$$

which is commonly seen in Calculus textbook.

Example 2.51. Let $f : \mathbb{R} \to \mathbb{R}$ and $F : \mathbb{R}^2 \to \mathbb{R}$ be differentiable, and $F\big(x, f(x)\big) = 0$ and $\dfrac{\partial F}{\partial y} \neq 0$. Then $f'(x) = -\dfrac{F_x\big(x, f(x)\big)}{F_y\big(x, f(x)\big)}$, where $F_x = \dfrac{\partial F}{\partial x}$ and $F_y = \dfrac{\partial F}{\partial y}$.

Example 2.52. Let $\gamma : (0, 1) \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Let $F(t) = f\big(\gamma(t)\big)$. Then $F'(t) = (Df)\big(\gamma(t)\big)\gamma'(t)$.

Example 2.53. Let $f(u, v, w) = u^2v + wv^2$ and $g(x, y) = (xy, \sin x, e^x)$. Let $h = f \circ g : \mathbb{R}^2 \to \mathbb{R}$. Find $\dfrac{\partial h}{\partial x}$.

Way I: Compute $\dfrac{\partial h}{\partial x}$ directly: Since

$$h(x, y) = f(g(x, y)) = f(xy, \sin x, e^x) = x^2y^2 \sin x + e^x \sin^2 x\,,$$

we have

$$\frac{\partial h}{\partial x} = 2xy^2 \sin x + x^2y^2 \cos x + e^x \sin^2 x + 2e^x \sin x \cos x\,.$$

Way II: Use the chain rule:

$$\begin{aligned} \frac{\partial h}{\partial x} &= \frac{\partial f}{\partial u}\frac{\partial g_1}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial g_2}{\partial x} + \frac{\partial f}{\partial w}\frac{\partial g_3}{\partial x} = 2uv \cdot y + (u^2 + 2wv) \cdot \cos x + v^2 \cdot e^x \\ &= 2xy^2 \sin x + (x^2y^2 + 2e^x \sin x) \cos x + e^x \sin^2 x. \end{aligned}$$

Example 2.54. Let $F(x, y) = f(x^2 + y^2)$, $f : \mathbb{R} \to \mathbb{R}$, $F : \mathbb{R}^2 \to \mathbb{R}$. Show that $x\dfrac{\partial F}{\partial y} = y\dfrac{\partial F}{\partial x}$.

Proof: Let $g(x, y) = x^2 + y^2$, $g : \mathbb{R}^2 \to \mathbb{R}$, then $F(x, y) = (f \circ g)(x, y)$. By the chain rule,

$$\begin{bmatrix} \dfrac{\partial F}{\partial x} & \dfrac{\partial F}{\partial y} \end{bmatrix} = f'(g(x, y)) \cdot \begin{bmatrix} \dfrac{\partial g}{\partial x} & \dfrac{\partial g}{\partial y} \end{bmatrix} = f'(g(x, y)) \begin{bmatrix} 2x & 2y \end{bmatrix}$$

which implies that

$$\frac{\partial F}{\partial x} = 2xf'(g(x, y)), \quad \frac{\partial F}{\partial y} = 2yf'(g(x, y))\,.$$

So $y\dfrac{\partial F}{\partial x} = f'(g(x, y))2xy = x\dfrac{\partial F}{\partial y}$.

2.5.4 The Mean Value Theorem

Theorem 2.55. *Let* $\mathcal{U} \subseteq \mathbb{R}^n$ *be open, and* $f : \mathcal{U} \to \mathbb{R}^m$ *with* $f = (f_1, \cdots, f_m)$*. Suppose that* f *is differentiable on* $\mathcal{U}$ *and the line segment joining* x *and* y *lies in* $\mathcal{U}$*. Then there exist points* $c_1, \cdots, c_m$ *on that segment such that*

$$f_i(y) - f_i(x) = (Df_i)(c_i)(y - x) \qquad \forall\, i = 1, \cdots, m.$$

Moreover, if $\mathcal{U}$ *is convex and* $\sup\limits_{x \in \mathcal{U}} \|(Df)(x)\|_{\mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)} \leqslant M$*, then*

$$\|f(x) - f(y)\|_{\mathbb{R}^m} \leqslant M\|x - y\|_{\mathbb{R}^n} \qquad \forall\, x, y \in \mathcal{U}\,.$$

Proof. Let $\gamma : [0,1] \to \mathbb{R}^n$ be given by $\gamma(t) = (1-t)x + ty$. Then by Theorem 2.49, for each $i = 1, \cdots, m$, $(f_i \circ \gamma) : [0,1] \to \mathbb{R}$ is differentiable on $(0,1)$; thus the mean value theorem (Theorem A.9) implies that there exists $t_i \in (0,1)$ such that

$$f_i(y) - f_i(x) = (f_i \circ \gamma)(1) - (f_i \circ \gamma)(0) = (f_i \circ \gamma)'(t_i) = (Df_i)(c_i)\big(\gamma'(t_i)\big)\,,$$

where $c_i = \gamma(t_i)$. On the other hand, $\gamma'(t_i) = y - x$.

Let $g(t) = (f \circ \gamma)(t)$. Then the chain rule implies that $g'(t) = (Df)(\gamma(t))(y - x)$; thus

$$\|g'(t)\|_{\mathbb{R}^m} \leqslant \|(Df)(\gamma(t))\|_{\mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)} \|y - x\|_{\mathbb{R}^m} \leqslant M\|x - y\|_{\mathbb{R}^n}\,.$$

Define $h(t) = \big(g(1) - g(0)\big) \cdot g(t)$. Then $h : [0,1] \to \mathbb{R}$ is differentiable; thus by the mean value theorem (Theorem A.9) we find that there exists $\xi \in (0,1)$ such that

$$h(1) - h(0) = h'(\xi) = \big(g(1) - g(0)\big) \cdot g'(\xi)\,;$$

thus by the fact that $g(0) = f(x)$ and $g(1) = f(y)$,

$$\begin{aligned} \|f(x) - f(y)\|^2_{\mathbb{R}^m} &= h(1) - h(0) \leqslant \|g(1) - g(0)\|_{\mathbb{R}^m} \|g'(\xi)\|_{\mathbb{R}^m} \\ &\leqslant M\|f(x) - f(y)\|_{\mathbb{R}^m} \|x - y\|_{\mathbb{R}^n} \end{aligned}$$

which concludes the theorem. □

Example 2.56. Let $f : [0,1] \to \mathbb{R}^2$ be given by $f(t) = (t^2, t^3)$. Then there is no $s \in (0,1)$ such that

$$(1,1) = f(1) - f(0) = f'(s)(1 - 0) = f'(s)$$

since $f'(s) = (2s, 3s^2) \neq (1,1)$ for all $s \in (0,1)$.

Example 2.57. Let $f:\mathbb{R}\to\mathbb{R}^2$ be given by $f(x)=(\cos x,\sin x)$. Then $f(2\pi)-f(0)=(0,0)$; however, $f'(x)=(-\sin x,\cos x)$ which cannot be a zero vector.

Example 2.58. Let f be given in Example 2.32, and $\mathcal{U}$ be a small neighborhood of the curve

$$\mathcal{C}=\big\{(x,y)\,\big|\,x^2+y^2=1,x\leqslant 0\big\}\cup\big\{(x,\pm 1)\,|\,0\leqslant x\leqslant 1\big\}.$$

Then

$$f(1,-1)-f(1,1)=\frac{3\pi}{2}.$$

On the other hand,

$$(Df)(x,y)(0,-2)=\begin{bmatrix}\dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2}\end{bmatrix}\begin{bmatrix}0\\-2\end{bmatrix}=-\frac{2x}{x^2+y^2}$$

which can never be $\dfrac{3\pi}{2}$ since $\Big|\dfrac{2x}{x^2+y^2}\Big|\leqslant 3$ if $(x,y)\in\mathcal{U}$ while $\dfrac{3\pi}{2}>3$. Therefore, no point (x,y) in $\mathcal{U}$ validates

$$(Df)(x,y)\big((1,-1)-(1,1)\big)=f(1,-1)-f(1,1).$$

Example 2.59. Suppose that $\mathcal{U}\subseteq\mathbb{R}^n$ is an open convex set, and $f:\mathcal{U}\to\mathbb{R}^m$ is differentiable and $Df(x)=0$ for all $x\in\mathcal{U}$. Then f is a constant; that is, for some $\alpha\in\mathbb{R}^m$ we have $f(x)=\alpha$ for all $x\in\mathcal{U}$.
Reason: Since $\mathcal{U}$ is convex, then the Mean Value Theorem can be applied to any $x,y\in\mathcal{U}$ such that $f_i(x)-f_i(y)=Df_i(c_i)(x-y)=0$ ($\because Df_i=0$) for $i=1,2,\cdots,m$; thus $f(x)=f(y)$ for any $x,y\in\mathcal{U}$. Let $\alpha=f(x)\in\mathbb{R}^m$, then we reach the conclusion.

2.6 The Inverse Function Theorem（反函數定理）

反函數定理是用來探討一個函數的反函數是否存在的問題。只要一個函數不是一對一的，一般來說都不能定義其反函數，例如三角函數中，正弦、餘弦及正切函數都是周期函數，所以全域的反函數不存在。但是我們也知道有所謂的反三角函數 $\sin^{-1}$（或 arcsin），$\cos^{-1}$（或 arctan）及 $\tan^{-1}$（或 arctan），這是因為我們限制了原三角函數的定義域使其在新的定義域上是一對一的（因此反函數存在）。因此，要討論一個定在某一個（大範圍的）定義域的函數的反函數，常常我們最多只能說反函數只在某一小塊區域上存在。

如何知道一個函數在一小塊區域上的反函數存在，我們首先該問的是在定義域是一維（或是指單變數函數）的情況下發生什麼事？由一維的反函數定理 (Theorem A.10) 我們知

道首先應該要保留的條件是類似於微分不為零的這個條件。但是在多變數函數之下，微分不為零的條件該怎麼呈現，這是第一個問題。而當我們觀察 (A.1)，應該可以猜出在多變數版本裡面所該對應到的條件，即是 $(Df)(x)$ 這個 bounded linear map 的可逆性。

另外，假設 $f \in \mathscr{C}^1$，那麼由 Theorem 1.87 我們知道在一個點 x_0 如果 $(Df)(x_0)$ 可逆的話，那麼在一個鄰域裡 $(Df)(x)$ 都可逆。所以下面這個反函數定理的條件中只有 (Df) 在一個點可逆這個條件，因為我們暫時也只能討論在小區域的反函數存不存在。

Before proceeding, we first prove the following important proposition which is used crucially in the proof of the inverse function theorem.

Proposition 2.60 (Contraction Mapping Principle)**.** *Let $F \subseteq \mathbb{R}^n$ be a closed subset (on which every Cauchy sequence converges), and $\Phi : F \to F$ be a **contraction mapping**; that is, there is a constant $\theta \in [0,1)$ such that*

$$\big\|\Phi(x) - \Phi(y)\big\|_{\mathbb{R}^n} \leqslant \theta\|x-y\|_{\mathbb{R}^n}\,.$$

*Then there exists a unique point $x \in F$, called the **fixed-point** of Φ, such that $\Phi(x) = x$.*

Proof. Let $x_0 \in F$, and define $x_{k+1} = \Phi(x_k)$ for all $k \in \mathbb{N} \cup \{0\}$. Then

$$\|x_{k+1} - x_k\|_{\mathbb{R}^n} = \big\|\Phi(x_k) - \Phi(x_{k-1})\big\|_{\mathbb{R}^n} \leqslant \theta\|x_k - x_{k-1}\|_{\mathbb{R}^n} \leqslant \cdots \leqslant \theta^k\|x_1 - x_0\|_{\mathbb{R}^n}\,;$$

thus if $\ell > k$,

$$\begin{aligned}\|x_\ell - x_k\|_{\mathbb{R}^n} &\leqslant \|x_k - x_{k+1}\|_{\mathbb{R}^n} + \|x_{k+1} - x_{k+2}\|_{\mathbb{R}^n} + \cdots + \|x_{\ell-1} - x_\ell\|_{\mathbb{R}^n} \\ &\leqslant (\theta^k + \theta^{k+1} + \cdots + \theta^{\ell-1})\|x_1 - x_0\|_{\mathbb{R}^n} \\ &\leqslant \theta^k(1 + \theta + \theta^2 + \cdots)\|x_1 - x_0\|_{\mathbb{R}^n} = \frac{\theta^k}{1-\theta}\|x_1 - x_0\|_{\mathbb{R}^n}\,. \end{aligned} \tag{2.8}$$

Since $\theta \in [0,1)$, $\lim\limits_{k\to\infty} \dfrac{\theta^k}{1-\theta}\|x_1 - x_0\|_{\mathbb{R}^n} = 0$; thus

$$\forall\, \varepsilon > 0, \exists\, N > 0 \ni \|x_k - x_\ell\|_{\mathbb{R}^n} < \varepsilon \quad \forall\, k, \ell \geqslant N\,.$$

In other words, $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence in F. By assumption, $x_k \to x$ as $k \to \infty$ for some $x \in F$. Finally, since $\Phi(x_k) = x_{k+1}$ for all $k \in \mathbb{N}$, by the continuity of Φ we obtain that

$$\Phi(x) = \lim_{k\to\infty} \Phi(x_k) = \lim_{k\to\infty} x_{k+1} = x$$

which guarantees the existence of a fixed-point.

Suppose that for some $x, y \in M$, $\Phi(x) = x$ and $\Phi(y) = y$. Then

$$\|x - y\|_{\mathbb{R}^n} = \big\|\Phi(x) - \Phi(y)\big\|_{\mathbb{R}^n} \leqslant \theta\|x - y\|_{\mathbb{R}^n}$$

which suggests that $\|x - y\|_{\mathbb{R}^n} = 0$ or $x = y$. Therefore, the fixed-point of Φ is unique. □

Now we state and prove the inverse function theorem.

Theorem 2.61 (Inverse Function Theorem)**.** *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open, $x_0 \in \mathcal{D}$, $f : \mathcal{D} \to \mathbb{R}^n$ be of class $\mathscr{C}^1$, and $(Df)(x_0)$ be invertible. Then there exist an open neighborhood $\mathcal{U}$ of x_0 and an open neighborhood $\mathcal{V}$ of $f(x_0)$ such that*

1. *$f : \mathcal{U} \to \mathcal{V}$ is one-to-one and onto;*
2. *The inverse function $f^{-1} : \mathcal{V} \to \mathcal{U}$ is of class $\mathscr{C}^1$;*
3. *If $x = f^{-1}(y)$, then $(Df^{-1})(y) = \big((Df)(x)\big)^{-1}$;*
4. *If f is of class $\mathscr{C}^r$ for some $r > 1$, so is f^{-1}.*

Proof. We will omit the proof of 4 since it requires more knowledge about differentiation.

Assume that $A = (Df)(x_0)$. Then $\|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} \neq 0$. Choose $\lambda > 0$ such that $2\lambda\|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} = 1$. Since $f \in \mathscr{C}^1$, there exists $\delta > 0$ such that

$$\big\|(Df)(x) - A\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} = \big\|(Df)(x) - (Df)(x_0)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} < \lambda \quad \text{whenever } x \in B(x_0, \delta) \cap \mathcal{D}\,.$$

By choosing δ even smaller if necessary, we can assume that $B(x_0, \delta) \subseteq \mathcal{D}$. Let $\mathcal{U} = B(x_0, \delta)$.
Claim: $f : \mathcal{U} \to \mathbb{R}^n$ is one-to-one (hence $f : \mathcal{U} \to f(\mathcal{U})$ is one-to-one and onto).
Proof of claim: For each $y \in \mathbb{R}^n$, define $\varphi_y(x) = x + A^{-1}\big(y - f(x)\big)$ (and we note that every fixed-point of φ_y corresponds to a solution to $f(x) = y$). Then

$$(D\varphi_y)(x) = \mathrm{Id} - A^{-1}(Df)(x) = A^{-1}\big(A - (Df)(x)\big)\,,$$

where Id is the identity map on $\mathbb{R}^n$. Therefore,

$$\big\|(D\varphi_y)(x)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} \leqslant \|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)}\big\|A - (Df)(x)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} < \frac{1}{2} \quad \forall\, x \in B(x_0, \delta)\,.$$

By the mean value theorem (Theorem 2.55),

$$\big\|\varphi_y(x_1) - \varphi_y(x_2)\big\|_{\mathbb{R}^n} \leqslant \frac{1}{2}\|x_1 - x_2\|_{\mathbb{R}^n} \qquad \forall\, x_1, x_2 \in B(x_0, \delta),\ x_1 \neq x_2\,; \tag{2.9}$$

thus at most one x satisfies $\varphi_y(x) = x$; that is, φ_y has at most one fixed-point. As a consequence, $f : B(x_0, \delta) \to \mathbb{R}^n$ is one-to-one.

Claim: The set $\mathcal{V} = f(\mathcal{U})$ is open.

Proof of claim: Let $b \in \mathcal{V}$. Then there is $a \in \mathcal{V}$ with $f(a) = b$. Choose $r > 0$ such that $\overline{B(a,r)} \subseteq \mathcal{U}$. We observe that if $y \in B(b, \lambda r)$, then

$$\|\varphi_y(a) - a\|_{\mathbb{R}^n} \leqslant \|A^{-1}\big(y - f(a)\big)\|_{\mathbb{R}^n} \leqslant \|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)}\|y - b\|_{\mathbb{R}^n} < \lambda\|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} r = \frac{r}{2};$$

thus if $y \in B(b, \lambda r)$ and $x \in B(a, r)$,

$$\|\varphi_y(x) - a\|_{\mathbb{R}^n} \leqslant \big\|\varphi_y(x) - \varphi_y(a)\big\|_{\mathbb{R}^n} + \|\varphi_y(a) - a\|_{\mathbb{R}^n} < \frac{1}{2}\|x - a\|_{\mathbb{R}^n} + \frac{r}{2} < r.$$

Therefore, if $y \in B(b, \lambda r)$, then $\varphi_y : B(a, r) \to B(a, r)$. By the continuity of φ_y,

$$\varphi_y : \overline{B(a,r)} \to \overline{B(a,r)}.$$

On the other hand, (2.9) implies that φ_y is a contraction mapping if $y \in B(b, \lambda r)$; thus by the contraction mapping principle (Proposition 2.60) φ_y has a unique fixed-point $x \in B(a, r)$. As a result, every $y \in B(b, \lambda r)$ corresponds to a unique $x \in B(a, r)$ such that $\varphi_y(x) = x$ or equivalently, $f(x) = y$. Therefore,

$$B(b, \lambda r) \subseteq f\big(B(a, r)\big) \subseteq f(\mathcal{U}) = \mathcal{V}.$$

Next we show that $f^{-1} : \mathcal{V} \to \mathcal{U}$ is differentiable. We note that if $x \in B(x_0, \delta)$,

$$\|(Df)(x_0) - (Df)(x)\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)}\|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} < \lambda\|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} = \frac{1}{2};$$

thus Theorem 1.87 implies that $(Df)(x)$ is invertible if $x \in B(x_0, \delta)$.

Let $b \in \mathcal{V}$ and $k \in \mathbb{R}^n$ such that $b + k \in \mathcal{V}$. Then there exists a unique $a \in \mathcal{U}$ and $h = h(k) \in \mathbb{R}^n$ such that $a + h \in \mathcal{U}$, $b = f(a)$ and $b + k = f(a + h)$. By the mean value theorem and (2.9),

$$\big\|\varphi_y(a + h) - \varphi_y(a)\big\|_{\mathbb{R}^n} < \frac{1}{2}\|h\|_{\mathbb{R}^n};$$

thus the fact that $f(a + h) - f(a) = k$ implies that

$$\|h - A^{-1}k\|_{\mathbb{R}^n} < \frac{1}{2}\|h\|_{\mathbb{R}^n}$$

which further suggests that

$$\frac{1}{2}\|h\|_{\mathbb{R}^n} \leqslant \|A^{-1}k\|_{\mathbb{R}^n} \leqslant \|A^{-1}\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)}\|k\|_{\mathbb{R}^n} \leqslant \frac{1}{2\lambda}\|k\|_{\mathbb{R}^n}. \tag{2.10}$$

As a consequence, if k is such that $b+k\in\mathcal{V}$,

$$\begin{aligned}
\frac{\left\|f^{-1}(b+k)-f^{-1}(b)-\big((Df)(a)\big)^{-1}k\right\|_{\mathbb{R}^n}}{\|k\|_{\mathbb{R}^n}} &= \frac{\left\|a+h-a-\big((Df)(a)\big)^{-1}k\right\|_{\mathbb{R}^n}}{\|k\|_{\mathbb{R}^n}} \\
&\leqslant \left\|\big((Df)(a)\big)^{-1}\right\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} \frac{\left\|k-(Df)(a)(h)\right\|_{\mathbb{R}^n}}{\|k\|_{\mathbb{R}^n}} \\
&\leqslant \left\|\big((Df)(a)\big)^{-1}\right\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)} \frac{\left\|f(a+h)-f(a)-(Df)(a)(h)\right\|_{\mathbb{R}^n}}{\|h\|_{\mathbb{R}^n}} \frac{\|h\|_{\mathbb{R}^n}}{\|k\|_{\mathbb{R}^n}} \\
&\leqslant \frac{\left\|\big((Df)(a)\big)^{-1}\right\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)}}{\lambda} \frac{\left\|f(a+h)-f(a)-(Df)(a)(h)\right\|_{\mathbb{R}^n}}{\|h\|_{\mathbb{R}^n}} .
\end{aligned}$$

Using (2.10), $h\to 0$ as $k\to 0$; thus passing $k\to 0$ on the left-hand side of the inequality above, by the differentiability of f we conclude that

$$\lim_{k\to 0}\frac{\left\|f^{-1}(b+k)-f^{-1}(b)-\big((Df)(a)\big)^{-1}k\right\|_{\mathbb{R}^n}}{\|k\|_{\mathbb{R}^n}}=0 .$$

This proves 3. □

Remark 2.62. Since $f^{-1}:\mathcal{V}\to\mathcal{U}$ is continuous, for any open subset $\mathcal{W}$ of $\mathcal{U}$ $f(\mathcal{W})=(f^{-1})^{-1}(\mathcal{W})$ is open relative to $\mathcal{V}$, or $f(\mathcal{W})=\mathcal{O}\cap\mathcal{V}$ for some open set $\mathcal{O}\subseteq\mathbb{R}^n$. In other words, if $\mathcal{U}$ is an open neighborhood of x_0 given by the inverse function theorem, then $f(\mathcal{W})$ is also open for all open subsets $\mathcal{W}$ of $\mathcal{U}$. We call this property as f is a ***local open mapping*** at x_0.

Remark 2.63. Since $(Df)(x_0)\in\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)$, the condition that $(Df)(x_0)$ is invertible can be replaced by that the determinant of the Jacobian matrix of f at x_0 is not zero; that is,

$$\det\big(\big[(Df)(x_0)\big]\big)\neq 0 .$$

The determinant of the Jacobian matrix of f at x_0 is called the ***Jacobian*** of f at x_0. The Jacobian of f at x sometimes is denoted by $\mathbf{J}_f(x)$ or $\dfrac{\partial(f_1,\cdots,f_n)}{\partial(x_1,\cdots,x_n)}$.

Example 2.64. Let $f:\mathbb{R}\to\mathbb{R}$ be given by

$$f(x)=\left\{\begin{array}{cl} x+2x^2\sin\dfrac{1}{x} & \text{if } x\neq 0 , \\ 0 & \text{if } x=0 . \end{array}\right.$$

Let $0\in(a,b)$ for some (small) open interval (a,b). Since $f'(x)=1-2\cos\dfrac{1}{x}+4x\sin\dfrac{1}{x}$ for $x\neq 0$, f has infinitely many critical points in (a,b), and (for whatever reasons) these critical points are local maximum points or local minimum points of f which implies that f is **not** locally invertible even though we have $f'(0) = 1 \neq 0$. One cannot apply the inverse function theorem in this case since f is not $\mathscr{C}^1$.

Corollary 2.65. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $f : \mathcal{U} \to \mathbb{R}^n$ be of class $\mathscr{C}^1$, and $(Df)(x)$ be invertible for all $x \in \mathcal{U}$. Then $f(\mathcal{W})$ is open for every open set $\mathcal{W} \subseteq \mathcal{U}$.*

在證明了小區域的（local）反函數定理 (Theorem 2.61) 之後，我們接下來要問的是全域的（global）反函數在什麼條件之下會存在。如果照一維的反函數定理，我們會猜測是不是只要 $(Df)(x)$ 在整個區域都可逆就能得到在全域的反函數都存在。以下給個反例說單單在這個條件之下，函數不一定會有一對一的性質。

Example 2.66. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f(x,y) = (e^x \cos y, e^x \sin y).$$

Then

$$\big[(Df)(x,y)\big] = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}.$$

It is easy to see that the Jacobian of f at any point is not zero (thus $(Df)(x)$ is invertible for all $x \in \mathbb{R}^2$), and f is not globally one-to-one (thus the inverse of f does not exist globally) since for example, $f(x,y) = f(x, y + 2\pi)$.

要再加什麼條件進來才能得到反函數在全域都存在是個不容易的問題。在一維的情況下，導數是 sign definite 就表示函數在全域是嚴格單調的。在高維度的情況，即使是 $(Df)(x)$ 到處都可逆，仍然有很多情況可能發生（如上例）。下面這個定理（全域的反函數存在定理），從某種角度來說並沒有真的加了什麼條件以確保全域的反函數存在，只是多要求了在所考慮的區域邊界上函數是一對一的。這個條件在一維的情況之下是自動成立的：因為如果一單變數函數的導數是 sign definite，那麼函數在邊界上必定是一對一的（因為嚴格單調的關係）。

Theorem 2.67 (Global Existence of Inverse Function)**.** *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open, $f : \mathcal{D} \to \mathbb{R}^n$ be of class $\mathscr{C}^1$, and $(Df)(x)$ be invertible for all $x \in K$. Suppose that K is a connected*（連通，即只有一塊）*, closed and bounded subset of $\mathcal{D}$, and $f : \partial K \to \mathbb{R}^n$ is one-to-one. Then $f : K \to \mathbb{R}^n$ is one-to-one.*

全域的反函數定理的證明需要更多關於點拓的知識，所以不在這門課中證明。

2.7 The Implicit Function Theorem（隱函數定理）

Theorem 2.68 (Implicit Function Theorem)**.** *Let $\mathcal{D} \subseteq \mathbb{R}^{\mathrm{n}} \times \mathbb{R}^{\mathrm{m}}$ be open, and $F : \mathcal{D} \to \mathbb{R}^{\mathrm{m}}$ be a function of class $\mathscr{C}^1$. Suppose that for some $(x_0, y_0) \in \mathcal{D}$, where $x_0 \in \mathbb{R}^{\mathrm{n}}$ and $y_0 \in \mathbb{R}^{\mathrm{m}}$, $F(x_0, y_0) = 0$ and*

$$\big[(D_yF)(x_0, y_0)\big] = \begin{bmatrix} \dfrac{\partial F_1}{\partial y_1} & \cdots & \dfrac{\partial F_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial y_1} & \cdots & \dfrac{\partial F_m}{\partial y_m} \end{bmatrix} (x_0, y_0)$$

is invertible. Then there exists an open neighborhood $\mathcal{U} \subseteq \mathbb{R}^{\mathrm{n}}$ of x_0, an open neighborhood $\mathcal{V} \subseteq \mathbb{R}^{\mathrm{m}}$ of y_0, and $f : \mathcal{U} \to \mathcal{V}$ such that

1. *$F\big(x, f(x)\big) = 0$ for all $x \in \mathcal{U}$;*
2. *$y_0 = f(x_0)$;*
3. *$(Df)(x) = -\big((D_yF)(x, f(x))\big)^{-1}(D_xF)\big(x, f(x)\big)$ for all $x \in \mathcal{U}$;*
4. *f is of class $\mathscr{C}^1$;*
5. *If F is of class $\mathscr{C}^r$ for some $r > 1$, so is f.*

Proof. Let $z = (x, y)$ and $w = (u, v)$, where $x, u \in \mathbb{R}^{\mathrm{n}}$ and $y, v \in \mathbb{R}^{\mathrm{m}}$. Define $w = G(z)$, where G is given by $G(x, y) = \big(x, F(x, y)\big)$. Then $G : \mathcal{D} \to \mathbb{R}^{n+m}$, and

$$\big[(DG)(x, y)\big] = \begin{bmatrix} \mathbb{I}_n & 0 \\ (D_xF)(x, y) & (D_yF)(x, y) \end{bmatrix},$$

where $\mathbb{I}_n$ is the $n \times n$ identity matrix and $(D_xF)(x, y) \in \mathscr{B}(\mathbb{R}^{\mathrm{n}}, \mathbb{R}^{\mathrm{m}})$ whose matrix representation is given by

$$\big[(D_xF)(x, y)\big] = \begin{bmatrix} \dfrac{\partial F_1}{\partial x_1} & \cdots & \dfrac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_m}{\partial x_1} & \cdots & \dfrac{\partial F_m}{\partial x_n} \end{bmatrix} (x, y)\,.$$

We note that the Jacobian of G at (x_0, y_0) is $\det\big([(D_yF)(x_0, y_0)]\big)$ which does not vanish since $(D_yF)(x_0, y_0)$ is invertible, so the inverse function theorem implies that there exists open neighborhoods $\mathcal{O}$ of (x_0, y_0) and $\mathcal{W}$ of $\big(x_0, F(x_0, y_0)\big) = (x_0, 0)$ such that

(a) $G : \mathcal{O} \to \mathcal{W}$ is one-to-one and onto;

(b) the inverse function $G^{-1} : \mathcal{W} \to \mathcal{O}$ is of class $\mathscr{C}^r$;

(c) $(DG^{-1})\big(x, F(x,y)\big) = \big((DG)(x,y)\big)^{-1}$.

By Remark 2.62, W.L.O.G. we can assume that $\mathcal{O} = \mathcal{U} \times \mathcal{V}$, where $\mathcal{U} \subseteq \mathbb{R}^n$ and $\mathcal{V} \subseteq \mathbb{R}^m$ are open, and $x_0 \in \mathcal{U}$, $y_0 \in \mathcal{V}$.

Write $G^{-1}(u,v) = \big(\varphi(u,v), \psi(u,v)\big)$, where $\varphi : \mathcal{W} \to \mathcal{U}$ and $\psi : \mathcal{W} \to \mathcal{V}$. Then

$$(u,v) = G\big(\varphi(u,v), \psi(u,v)\big) = \big(\varphi(u,v), F(u, \psi(u,v))\big)$$

which implies that $\varphi(u,v) = u$ and $v = F(u, \psi(u,v))$. Let $f(x) = \psi(x,0)$. Then $\big(u, f(u)\big) \in \mathcal{U} \times \mathcal{V}$ is the unique point satisfying $F\big(u, f(u)\big) = 0$ if $u \in \mathcal{U}$. Therefore, $f : \mathcal{U} \to \mathcal{V}$, and

$$F\big(x, f(x)\big) = 0 \qquad \forall\, x \in \mathcal{U}.$$

Since $G(x_0, y_0) = (x_0, 0) = G\big(x_0, f(x_0)\big)$, $(x_0, y_0), \big(x_0, f(x_0)\big) \in \mathcal{O}$, and $G : \mathcal{O} \to \mathcal{W}$ is one-to-one, we must have $y_0 = f(x_0)$.

By (b) and (c), we have G^{-1} is of class $\mathscr{C}^1$, and

$$(DG^{-1})(u,v) = \big((DG)(x,y)\big)^{-1}.$$

As a consequence, $\psi \in \mathscr{C}^1$, and

$$\begin{aligned}\begin{bmatrix} (D_u\varphi)(u,v) & (D_v\varphi)(u,v) \\ (D_u\psi)(u,v) & (D_v\psi)(u,v) \end{bmatrix} &= \begin{bmatrix} \mathbb{I}_n & 0 \\ (D_xF)(x,y) & (D_yF)(x,y) \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbb{I}_n & 0 \\ -\big((D_yF)(x,y)\big)^{-1}(D_xF)(x,y) & \big((D_yF)(x,y)\big)^{-1} \end{bmatrix}.\end{aligned}$$

Evaluating the equation above at $v = 0$, we conclude that

$$(Df)(u) = (D_u\psi)(u,0) = -\big((D_yF)(u, f(u))\big)^{-1}(D_xF)\big(u, f(u)\big)$$

which implies 3. We also note that 4 follows from (b) and 5 follows from 3. □

Example 2.69. Let $F(x,y) = x^2 + y^2 - 1$.

1. If $(x_0, y_0) = (1,0)$, then $F_x(x_0, y_0) = 2 \neq 0$; thus the implicit function theorem implies that locally x can be expressed as a function of y.

2. If $(x_0, y_0) = (0, -1)$, then $F_y(x_0, y_0) = -2 \neq 0$; thus the implicit function theorem implies that locally y can be expressed as a function of x.

3. If $(x_0, y_0) = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$, then $F_x(x_0, y_0) = -1 \neq 0$ and $F_y(x_0, y_0) = \sqrt{3} \neq 0$; thus the implicit function theorem implies that locally x can be expressed as a function of y and locally y can be expressed as a function of x.

Example 2.70. Suppose that (x, y, u, v) satisfies the equation

$$\left\{\begin{array}{r} xu + yv^2 = 0 \\ xv^3 + y^2u^6 = 0 \end{array}\right.$$

and $(x_0, y_0, u_0, v_0) = (1, -1, 1, -1)$. Let $F(x, y, u, v) = (xu + yv^2, xv^3 + y^2u^6)$. Then $F(x_0, y_0, u_0, v_0) = 0$.

1. Since $(D_{x,y}F)(x_0, y_0, u_0, v_0) = \begin{bmatrix} \dfrac{\partial F_1}{\partial x} & \dfrac{\partial F_1}{\partial y} \\ \dfrac{\partial F_2}{\partial x} & \dfrac{\partial F_2}{\partial y} \end{bmatrix}(x_0, y_0, u_0, v_0) = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$ is invertible, locally (x, y) can be expressed in terms of u, v; that is, locally $x = x(u, v)$ and $y = y(u, v)$.

2. Since $(D_{y,u}F)(x_0, y_0, u_0, v_0) = \begin{bmatrix} \dfrac{\partial F_1}{\partial y} & \dfrac{\partial F_1}{\partial u} \\ \dfrac{\partial F_2}{\partial y} & \dfrac{\partial F_2}{\partial u} \end{bmatrix}(x_0, y_0, u_0, v_0) = \begin{bmatrix} 1 & 1 \\ -2 & 6 \end{bmatrix}$ is invertible, locally (y, u) can be expressed in terms of x, v.

Example 2.71. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be given by

$$f(x, y, z) = (xe^y + ye^z, xe^z + ze^y).$$

Then f is of class $\mathscr{C}^1$, $f(-1, 1, 1) = (0, 0)$ and

$$\big[(Df)(x, y, z)\big] = \begin{bmatrix} e^y & xe^y + e^z & ye^z \\ e^z & ze^y & xe^z + e^y \end{bmatrix}.$$

Since $(D_{y,z}f)(-1, 1, 1) = \begin{bmatrix} 0 & e \\ e & 0 \end{bmatrix}$ is invertible, the implicit function theorem implies that the system

$$\left\{\begin{array}{l} xe^y + ye^z = 0 \\ xe^z + ze^y = 0 \end{array}\right.$$

can be solved for y and z as continuously differentiable function of x for x near -1 and (y,z) near $(1,1)$. Furthermore, if we write $(y,z) = g(x)$ for x near -1, then

$$g'(x) = \begin{bmatrix} xe^y + e^z & ye^z & ye^z \\ ze^y & xe^z + e^y & \end{bmatrix}^{-1} \begin{bmatrix} e^y \\ e^z \end{bmatrix}.$$

2.8 Directional Derivatives and Gradient Vectors

Definition 2.72 (Directional Derivatives)**.** Let f be real-valued and defined on a neighborhood of $x_0 \in \mathbb{R}^n$, and let $\mathrm{v} \in \mathbb{R}^n$ be a unit vector. Then

$$(D_{\mathrm{v}}f)(x_0) \equiv \frac{d}{dt}\Big|_{t=0} f(x_0 + t\mathrm{v}) = \lim_{t\to 0} \frac{f(x_0 + t\mathrm{v}) - f(x_0)}{t}$$

is called the ***directional derivative***（方向導數）of f at x_0 in the direction v.

Remark 2.73. Let $\{\mathrm{e}_j\}_{j=1}^n$ be the standard basis of $\mathbb{R}^n$. Then the partial derivative $\dfrac{\partial f}{\partial x_j}(x_0)$ (if it exists) is the directional derivative of f at x_0 in the direction e_j.

Remark 2.74. Let f be a real-valued differentiable function defined on a neighborhood of $x_0 \in \mathbb{R}^n$, and let $\mathrm{v} \in \mathbb{R}^n$ be a unit vector. For a curve $\gamma : (-\delta, \delta) \to \mathbb{R}^n$ satisfying that $\gamma(0) = x_0$ and $\gamma'(0) = \mathrm{v}$, the chain rule shows that

$$\frac{d}{dt}\Big|_{t=0} (f \circ \gamma)(t) = (Df)(x_0)(\mathrm{v}) = (D_{\mathrm{v}}f)(x_0)\,.$$

In other words, for a differentiable function f in a neighborhood of x_0, the derivative $\dfrac{d}{dt}\Big|_{t=0}(f \circ \gamma)$ is independent of γ as long as $\gamma(0) = x_0$ and $\gamma'(0) = \mathrm{v}$. Therefore, directional derivative of a differential function f at x_0 in the direction v can also be defined by the value $\dfrac{d}{dt}\Big|_{t=0}(f \circ \gamma)(t)$, where $\gamma : (-\delta, \delta) \to \mathbb{R}^n$ is any curve satisfying $\gamma(0) = x_0$ and $\gamma'(0) = \mathrm{v}$.

Theorem 2.75. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$ be differentiable at x_0. Then the directional derivative of f at x_0 in the direction* v *is $(Df)(x_0)(\mathrm{v})$.*

Proof. Since f is differentiable at x_0, $\forall\, \varepsilon > 0$, $\ni \delta > 0$ such that

$$\big|f(x) - f(x_0) - (Df)(x_0)(x - x_0)\big| \leqslant \frac{\varepsilon}{2}\|x - x_0\|_{\mathbb{R}^n} \text{ whenever } \|x - x_0\|_{\mathbb{R}^n} < \delta\,.$$

In particular, if $x = x_0 + t\mathrm{v}$ with v being a unit vector in $\mathbb{R}^{\mathrm{n}}$ and $0 < |t| < \delta$, then

$$\Big|\frac{f(x_0+t\mathrm{v})-f(x_0)}{t} - (Df)(x_0)(\mathrm{v})\Big| = \frac{\big|f(x_0+t\mathrm{v})-f(x_0)-(Df)(x_0)(t\mathrm{v})\big|}{|t|}$$
$$= \frac{\big|f(x)-f(x_0)-(Df)(x_0)(x-x_0)\big|}{|t|} \leqslant \frac{\varepsilon}{2} < \varepsilon\,;$$

thus $(D_{\mathrm{v}}f)(x_0) = (Df)(x_0)(\mathrm{v})$. ▫

Remark 2.76. When $v \in \mathbb{R}^{\mathrm{n}}$ but $0 < \|v\|_{\mathbb{R}^{\mathrm{n}}} \neq 1$, we let $\mathrm{v} = \dfrac{v}{\|v\|_{\mathbb{R}^{\mathrm{n}}}}$. Then the direction derivatives of a function $f : \mathcal{U} \subseteq \mathbb{R}^{\mathrm{n}} \to \mathbb{R}$ at $a \in \mathcal{U}$ in the direction v is

$$(D_{\mathrm{v}}f)(a) = \lim_{t\to 0}\frac{f(a+t\mathrm{v})-f(a)}{t}\,.$$

Making a change of variable $s = \dfrac{t}{\|v\|_{\mathbb{R}^{\mathrm{n}}}}$. Then

$$(Df)(x_0)(v) = \|v\|_{\mathbb{R}^{\mathrm{n}}}(Df)(x_0)(\mathrm{v}) = \|v\|_{\mathbb{R}^{\mathrm{n}}}\lim_{t\to 0}\frac{f(a+t\mathrm{v})-f(a)}{t} = \lim_{s\to 0}\frac{f(a+sv)-f(a)}{s}\,.$$

We sometimes also call the value $(Df)(x_0)(v)$ the "directional derivative" of f in the "direction" v.

Example 2.77. The existence of directional derivatives of a function f at x_0 in all directions does not guarantee the differentiability of f at x_0. For example, let $f : \mathbb{R}^2 \to \mathbb{R}$ be given as in Example 2.44, and $\mathrm{v} = (\mathrm{v}_1, \mathrm{v}_2) \in \mathbb{R}^2$ be a unit vector. Then

$$(D_{\mathrm{v}}f)(0) = \lim_{t\to 0}\frac{f(t\mathrm{v}_1, t\mathrm{v}_2)-f(0,0)}{t} = \mathrm{v}_1^3\,.$$

However, f is not differentiable at $(0,0)$. We also note that in this example, $(D_{\mathrm{v}}f)(0) \neq (Jf)(0)\mathrm{v}$, where $(Jf)(0) = \Big[\frac{\partial f}{\partial x}(0,0) \quad \frac{\partial f}{\partial y}(0,0)\Big]$ is the Jacobian matrix of f at $(0,0)$.

Example 2.78. The existence of directional derivatives of a function f at x_0 in all directions does not even guarantee the continuity of f at x_0. For example, let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4} & \text{if } (x,y) \neq (0,0)\,, \\ 0 & \text{if } (x,y) = (0,0)\,, \end{cases}$$

and $\mathrm{v} = (\mathrm{v}_1, \mathrm{v}_2) \in \mathbb{R}^2$ be a unit vector. Then if $\mathrm{v}_1 \neq 0$,

$$(D_{\mathrm{v}}f)(0) = \lim_{t\to 0}\frac{f(t\mathrm{v}_1, t\mathrm{v}_2)-f(0,0)}{t} = \lim_{t\to 0}\frac{t^3\mathrm{v}_1\mathrm{v}_2^2}{t(t^2\mathrm{v}_1^2+t^4\mathrm{v}_2^4)} = \frac{\mathrm{v}_2^2}{\mathrm{v}_1}$$

while if $v_1 = 0$,

$$(D_v f)(0) = \lim_{t\to 0} \frac{f(tv_1, tv_2) - f(0,0)}{t} = 0\,.$$

However, f is not continuous at $(0,0)$ since if (x, y) approaches $(0,0)$ along the curve $x = my^2$ with $m \neq 0$, we have

$$\lim_{y\to 0} f(my^2, y) = \lim_{y\to 0} \frac{my^4}{m^2y^4 + y^4} = \frac{m}{m^2+1}$$

which depends on m. Therefore, f is not continuous at $(0,0)$.

Example 2.79. Here comes another example showing that a function having directional derivative in all directions might not be continuous. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} \dfrac{xy}{x+y^2} & \text{if } x + y^2 \neq 0\,, \\ 0 & \text{if } x + y^2 = 0\,, \end{cases}$$

and $v = (v_1, v_2) \in \mathbb{R}^2$ be a unit vector. Then if $v_1 \neq 0$,

$$(D_v f)(0) = \lim_{t\to 0} \frac{f(tv_1, tv_2) - f(0,0)}{t} = \lim_{t\to 0} \frac{t^2 v_1 v_2}{t(tv_1 + t^2 v_2^2)} = v_2$$

while if $v_1 = 0$,

$$(D_v f)(0) = \lim_{t\to 0} \frac{f(tv_1, tv_2) - f(0,0)}{t} = 0\,.$$

However, f is not continuous at $(0,0)$ since if (x, y) approaches $(0,0)$ along the polar curve

$$\theta(r) = \frac{\pi}{2} + \sin^{-1}(r - mr^2) \qquad 0 < r \ll 1\,,$$

we have

$$\lim_{\substack{(x,y)\to(0,0)\\ x=r\cos\theta(r),\, y=r\sin\theta(r)}} f(x,y) = \lim_{r\to 0^+} \frac{r^2\cos\theta(r)\sin\theta(r)}{r^2\sin^2\theta(r) + r\cos\theta(r)} = \lim_{r\to 0^+} \frac{r(-r+mr^2)\sin\theta(r)}{r\sin^2\theta(r) - r + mr^2}$$

$$= \lim_{r\to 0^+} \frac{(-r+mr^2)\sin\theta(r)}{\sin^2\theta(r) - 1 + mr} = \frac{-1}{m}$$

which depends on m. Therefore, f is not continuous at $(0,0)$.

Definition 2.80. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be an open set. The derivative of a scalar function $f : \mathcal{U} \to \mathbb{R}$ is called the ***gradient*** of f and is denoted by $\text{grad} f$ or ∇f.

Let $\mathcal{U} \subseteq \mathbb{R}^{\rm n}$ be an open set, $a \in \mathcal{U}$ and $f : \mathcal{U} \to \mathbb{R}$ be a real-valued function. Suppose that $f \in \mathscr{C}^1(\mathcal{U};\mathbb{R})$ and $(\nabla f)(a) \neq 0$. Then $\dfrac{\partial f}{\partial x_k}(a) \neq 0$ for some $1 \leqslant k \leqslant {\rm n}$. W.L.O.G., we can assume that $\dfrac{\partial f}{\partial x_{\rm n}}(a) \neq 0$. By the implicit function theorem, there exists an open neighborhood $\mathcal{V} \subseteq \mathbb{R}^{{\rm n}-1}$ of $(a_1, \cdots, a_{{\rm n}-1})$ and an open neighborhood $\mathcal{W} \subseteq \mathbb{R}$ of $a_{\rm n}$, as well as a $\mathscr{C}^1$-function $\varphi : \mathcal{V} \to \mathbb{R}$ such that in a neighborhood of a the level set $\{x \in \mathcal{U} \,|\, f(x) = f(a)\}$ can be represented by $x_{\rm n} = \varphi(x_1, \cdots, x_{{\rm n}-1})$; that is,

$$f\big(x_1, \cdots, x_{{\rm n}-1}, \varphi(x_1, \cdots, x_{{\rm n}-1})\big) = f(a) \qquad \forall\,(x_1, \cdots, x_{{\rm n}-1}) \in \mathcal{V}.$$

Moreover,

$$\varphi_{x_j}(x_1, \cdots, x_{{\rm n}-1}) = -\frac{f_{x_j}\big(x_1, \cdots, x_{{\rm n}-1}, \varphi(x_1, \cdots, x_{{\rm n}-1})\big)}{f_{x_n}\big(x_1, \cdots, x_{{\rm n}-1}, \varphi(x_1, \cdots, x_{{\rm n}-1})\big)} \qquad \forall\,(x_1, \cdots, x_{{\rm n}-1}) \in \mathcal{V}.$$

Consider the collection of vectors $\{v_j\}_{j=1}^{{\rm n}-1}$ given by

$$v_j = \frac{\partial}{\partial x_j}\Big|_{x=a}\big(x_1, \cdots, x_{{\rm n}-1}, \varphi(x_1, \cdots, x_{{\rm n}-1})\big) \qquad (x_1, \cdots, x_{{\rm n}-1}) \in \mathcal{V}.$$

Then $v_j' s$ are tangent vectors of the level surface. If $\{{\rm e}_j\}_{j=1}^{\rm n}$ is the standard basis of $\mathbb{R}^{\rm n}$, then

$$v_j = e_j + \big(0, \cdots, 0, \varphi_{x_j}(a_1, \cdots, a_{{\rm n}-1})\big) = e_j - \Big(0, \cdots, 0, \frac{f_{x_j}(a)}{f_{x_n}(a)}\Big).$$

Therefore, the gradient vector $(\nabla f)(a)$ is perpendicular to v_j for all $1 \leqslant j \leqslant {\rm n}-1$ which conclude the following

Proposition 2.81. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open and $f \in \mathscr{C}^1(\mathcal{U};\mathbb{R})$; that is, $f : \mathcal{U} \to \mathbb{R}$ is continuously differentiable. Then if $(\nabla f)(x_0) \neq 0$, the vector $\dfrac{(\nabla f)(x_0)}{\|(\nabla f)(x_0)\|_{\mathbb{R}^n}}$ is the unit normal to the level set $\{x \in \mathcal{U} \,|\, f(x) = f(x_0)\}$ at x_0.*

Example 2.82. Find the normal to $\mathcal{S} = \{(x, y, z) \,|\, x^2 + y^2 + z^2 = 3\}$ at $(1, 1, 1) \in \mathcal{S}$.
Solution: Take $f(x, y, z) = x^2 + y^2 + z^2 - 3$. Then $(\nabla f)(x, y, z) = (2x, 2y, 2z)$; thus $(\nabla f)(1, 1, 1) = (2, 2, 2)$ is normal to S at $(1, 1, 1)$.

Example 2.83. Consider the surface

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 - y^2 + xyz = 1\}.$$

Find the tangent plane of $\mathcal{S}$ at $(1, 0, 1)$.

Solution: Let $f(x, y, z) = x^2 - y^2 + xyz$. Then

$$\mathcal{S} = \{(x, y, z) \in \mathbb{R}^3 \,|\, f(x, y, z) = f(1, 0, 1)\};$$

that is, $\mathcal{S}$ is a level set of f. Since $(\nabla f)(1, 0, 1) = (2, 1, 0) \neq (0, 0, 0)$, $(2, 1, 0)$ is normal to $\mathcal{S}$ at $(1, 0, 1)$; thus the tangent plane of S at $(1, 0, 1)$ is $2(x - 1) + y = 0$. □

Proposition 2.84. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. If $(\nabla f)(x_0) \neq 0$, then $\pm\dfrac{(\nabla f)(x_0)}{\|(\nabla f)(x_0)\|_{\mathbb{R}^n}}$ is the direction in which the function f increases/decreases most rapidly*（最速上升／下降方向）*at x_0.*

Proof. Let $x_0 \in \mathbb{R}^n$ be given. Suppose that f increases most rapidly in the direction v, then $(D_{\rm v} f)(x_0) = \sup\limits_{\|w\|_{\mathbb{R}^n}=1} (D_w f)(x_0)$. Since f is differentiable, $(D_w f)(x_0) = (Df)(x_0)(w) = (\nabla f)(x_0) \cdot w$ which is maximized in the direction $\dfrac{(\nabla f)(x_0)}{\|(\nabla f)(x_0)\|_{\mathbb{R}^n}}$. □

Example 2.85. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x, y, z) = x^2 y \sin z$. Find the direction of the greatest rate of change at $(3, 2, 0)$.

Solution: We compute the gradient of f at $(3, 2, 0)$ as follows:

$$\begin{aligned}(\nabla f)(3, 2, 0) &= \Big(\frac{\partial f}{\partial x}(3, 2, 0), \frac{\partial f}{\partial y}(3, 2, 0), \frac{\partial f}{\partial z}(3, 2, 0)\Big)\\ &= (2xy \sin z, x^2 \sin z, x^2 y \cos z)\big|_{(x,y,z)=(3,,2,0)} = (0, 0, 18).\end{aligned}$$

Therefore, the direction of the greatest rate of change of f at $(3, 2, 0)$ is $(0, 0, 1)$.