Chapter 6

Differentiation of Maps

6.1 Bounded Linear Maps

Definition 6.1. A map L from a vector space X into a vector space Y is said to be ***linear*** if $L(cx_1 + x_2) = cL(x_1) + L(x_2)$ for all $x_1, x_2 \in X$ and $c \in \mathbb{R}$. We often write Lx instead of $L(x)$, and the collection of all linear maps from X to Y is denoted by $\mathscr{L}(X, Y)$.

Suppose further that X and Y are normed spaces equipped with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. A linear map $L : X \to Y$ is said to be bounded if

$$\sup_{\|x\|_X=1} \|Lx\|_Y < \infty\,.$$

The collection of all bounded linear maps from X to Y is denoted by $\mathscr{B}(X, Y)$, and the number $\sup\limits_{\|x\|_X=1} \|Lx\|_Y$ is often denoted by $\|L\|_{\mathscr{B}(X,Y)}$.

Example 6.2. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be given by $Lx = Ax$, where A is an $m \times n$ matrix. Then Example 1.138 shows that $\|L\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)}$ is the square root of the largest eigenvalue of $A^{\mathrm{T}}A$ which is certainly a finite number. Therefore, any linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is bounded.

Proposition 6.3. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces, and $L \in \mathscr{B}(X, Y)$. Then*

$$\|L\|_{\mathscr{B}(X,Y)} = \sup_{x\neq 0} \frac{\|Lx\|_Y}{\|x\|_X} = \inf\big\{M > 0 \,\big|\, \|Lx\|_Y \leqslant M\|x\|_X\big\}\,.$$

In particular, the first equality implies that

$$\|Lx\|_Y \leqslant \|L\|_{\mathscr{B}(X,Y)}\|x\|_X \qquad \forall\, x \in X\,.$$

Proposition 6.4. *Let* $(X, \|\cdot\|_X)$ *and* $(Y, \|\cdot\|_Y)$ *be normed spaces, and* $L \in \mathscr{L}(X, Y)$*. Then* L *is continuous on* X *if and only if* $L \in \mathscr{B}(X, Y)$*.*

Proof. "$\Rightarrow$" Since L is continuous at $0 \in X$, there exists $\delta > 0$ such that

$$\|Lx\|_Y = \|Lx - L0\|_Y < 1 \quad \text{if} \ \ \|x\|_X < \delta\,.$$

Then $\big\|L\big(\frac{\delta}{2}x\big)\big\|_Y \leqslant 1$ if $\big\|\frac{\delta}{2}x\big\|_X < \delta$; thus by the properties of norm,

$$\|Lx\|_Y \leqslant \frac{2}{\delta} \qquad \text{if} \ \ \|x\|_X < 2\,.$$

Therefore, $\sup\limits_{\|x\|_X = 1} \|Lx\|_Y \leqslant \dfrac{2}{\delta}$ which implies that $L \in \mathscr{B}(X, Y)$.

"$\Leftarrow$" If $L \in \mathscr{B}(X, Y)$, then $M = \|L\|_{\mathscr{B}(X,Y)} < \infty$, and

$$\|Lx_1 - Lx_2\|_Y = \|L(x_1 - x_2)\|_Y \leqslant M\|x_1 - x_2\|_X$$

which shows that L is uniformly continuous on X. ▫

Proposition 6.5. *Let* $(X, \|\cdot\|_X)$ *and* $(Y, \|\cdot\|_Y)$ *be normed spaces. Then* $\big(\mathscr{B}(X, Y), \|\cdot\|_{\mathscr{B}(X,Y)}\big)$ *is a normed space. Moreover, if* $(Y, \|\cdot\|_Y)$ *is a Banach space, so is* $\big(\mathscr{B}(X, Y), \|\cdot\|_{\mathscr{B}(X,Y)}\big)$*.*

Proof. That $\big(\mathscr{B}(X, Y), \|\cdot\|_{\mathscr{B}(X,Y)}\big)$ is a normed space is left as an exercise. Now suppose that $\big(Y, \|\cdot\|_Y\big)$ is a Banach space. Let $\{L_k\}_{k=1}^{\infty} \subseteq \mathscr{B}(X, Y)$ be a Cauchy sequence. Then by Proposition 6.3, for each $x \in X$ we have

$$\|L_k x - L_\ell x\|_Y = \|(L_k - L_\ell)x\|_Y \leqslant \|L_k - L_\ell\|_{\mathscr{B}(X,Y)}\|x\|_X \to 0 \quad \text{as} \ \ k, \ell \to \infty\,.$$

Therefore, $\{L_k x\}_{k=1}^{\infty}$ is a Cauchy sequence in Y; thus convergent. Suppose that $\lim\limits_{k\to\infty} L_k x = y$. We then establish a map $x \mapsto y$ which we denoted by L; that is, $Lx = y$. Then L is linear since if $x_1, x_2 \in X$ and $c \in \mathbb{R}$,

$$L(cx_1 + x_2) = \lim_{k\to\infty} L_k(cx_1 + x_2) = \lim_{k\to\infty} \big(cL_k x_1 + L_k x_2\big) = cLx_1 + Lx_2\,.$$

Moreover, since $\{L_k\}_{k=1}^{\infty}$ is a Cauchy sequence, $\exists\, M > 0$ such that $\|L_k\|_{\mathscr{B}(X,Y)} \leqslant M$ for all $k \in \mathbb{N}$. If $\varepsilon > 0$ is given, for each $x \in X$ there exists $N = N_x > 0$ such that

$$\|L_k x - Lx\|_Y < \varepsilon \qquad \forall\, k \geqslant N_x\,.$$

Therefore, for $k \geqslant N_x$,

$$\|Lx\|_Y < \|L_k x\|_Y + \varepsilon \leqslant \|L_k\|_{\mathscr{B}(X,Y)}\|x\|_X + \varepsilon \leqslant M\|x\|_X + \varepsilon$$

which implies that $\sup\limits_{\|x\|_X=1} \|Lx\|_Y \leqslant M + \varepsilon$; thus $L \in \mathscr{B}(X,Y)$.

Finally, we show that $\lim\limits_{k\to\infty} \|L_k - L\|_{\mathscr{B}(X,Y)} = 0$. Let $x \in X$ and $\varepsilon > 0$ be given. Since $\{L_k\}_{k=1}^{\infty}$ is a Cauchy sequence, there exists $N > 0$ such that $\|L_k - L_\ell\|_{\mathscr{B}(X,Y)} < \dfrac{\varepsilon}{2}$ if $k, \ell \geqslant N$. Then if $k \geqslant N$,

$$\|L_k x - Lx\|_Y = \lim_{\ell\to\infty} \|L_k x - L_\ell x\|_Y \leqslant \limsup_{\ell\to\infty} \|L_k - L_\ell\|_{\mathscr{B}(X,Y)}\|x\|_X \leqslant \frac{\varepsilon}{2}\|x\|_X$$

which shows that $\|L_k - L\|_{\mathscr{B}(X,Y)} < \varepsilon$ if $k \geqslant N$. □

Proposition 6.6. *Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$, $(Z, \|\cdot\|_Z)$ be normed spaces, and $L \in \mathscr{B}(X,Y)$, $K \in \mathscr{B}(Y,Z)$. Then $K \circ L \in \mathscr{B}(X,Z)$, and*

$$\|K \circ L\|_{\mathscr{B}(X,Z)} \leqslant \|K\|_{\mathscr{B}(Y,Z)}\|L\|_{\mathscr{B}(X,Y)}\,.$$

We often write $K \circ L$ as KL if K and L are linear.

Proof. By the properties of the norm of a bounded linear map,

$$\|K \circ L(x)\|_Z = \|K(Lx)\|_Z \leqslant \|K\|_{\mathscr{B}(Y,Z)}\|Lx\|_Y \leqslant \|K\|_{\mathscr{B}(Y,Z)}\|L\|_{\mathscr{B}(X,Y)}\|x\|_X\,.$$ □

From now on, when the domain X and the target Y of a linear map L is clear, we use $\|L\|$ instead of $\|L\|_{\mathscr{B}(X,Y)}$ to simplify the notation.

Theorem 6.7. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces, and X be finite dimensional. Then every linear map from X to Y is bounded; that is, $\mathscr{L}(X,Y) = \mathscr{B}(X,Y)$.*

Proof. Suppose that $\dim(X) = n$. Let $\{\mathrm{e}_k\}_{k=1}^n \subseteq X$ be a linearly independent set of vectors. From Example 4.24, every two norms on X are equivalent; thus we only focus on the norm $\|\cdot\|_2$ on X induced by the inner product

$$\big(\mathrm{e}_i, \mathrm{e}_j\big)_X = \delta_{ij} \qquad \forall\, i = 1, \cdots n.$$

Since $\{\mathrm{e}_k\}_{k=1}^n$ is a linear independent set of vectors, every $x \in X$ can be expressed as a unique linear combination of e_k's; that is, for all $x \in X$, $\exists\, c_1 = c_1(x), \cdots, c_n = c_n(x) \in \mathbb{R}$ such that

$$x = c_1\mathrm{e}_1 + \cdots + c_n\mathrm{e}_n\,.$$

These coefficients c_k's in fact are determined by $c_k = (x, \mathrm{e}_k)_X$, and, by Example 4.24 and the Cauchy-Schwarz inequality, satisfy

$$\big|c_k(x)\big| \leqslant \|x\|_2 \|\mathrm{e}_k\|_2 \leqslant C\|x\|_X .$$

As a consequence, if L is a linear map from X to Y, then

$$\begin{aligned}\|Lx\|_Y &= \big\|L(c_1(x)\mathrm{e}_1 + \cdots c_n(x)\mathrm{e}_n)\big\|_Y \leqslant |c_1(x)|\|L\mathrm{e}_1\|_Y + \cdots |c_n(x)|\|L\mathrm{e}_n\|_Y \\ &\leqslant nC\|x\|_X \max\big\{\|L\mathrm{e}_1\|_Y, \cdots \|L\mathrm{e}_n\|_Y\big\} \leqslant M\|x\|_X\end{aligned}$$

for some constant $M > 0$; thus $\|L\|_{\mathscr{B}(X,Y)} \leqslant M < \infty$ which shows that $L \in \mathscr{B}(X, Y)$. □

Theorem 6.8. *Let* $\mathrm{GL}(n)$ *be the set of all invertible linear maps on* $\mathbb{R}^n$*; that is,*

$$\mathrm{GL}(n) = \big\{L \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n) \,\big|\, L \textit{ is one-to-one (and onto)}\big\} .$$

1. *If* $L \in \mathrm{GL}(n)$ *and* $K \in \mathscr{B}(\mathbb{R}^n, \mathbb{R}^n)$ *satisfying* $\|K - L\|\|L^{-1}\| < 1$, *then* $K \in \mathrm{GL}(n)$.
2. $\mathrm{GL}(n)$ *is an open set of* $\mathscr{B}(\mathbb{R}^n, \mathbb{R}^n)$.
3. *The mapping* $L \mapsto L^{-1}$ *is continuous on* $\mathrm{GL}(n)$.

Proof. 1. Let $\|L^{-1}\| = \dfrac{1}{\alpha}$ and $\|K - L\| = \beta$. Then $\beta < \alpha$; thus for every $x \in \mathbb{R}^n$,

$$\begin{aligned}\alpha\|x\|_{\mathbb{R}^n} &= \alpha\|L^{-1}Lx\|_{\mathbb{R}^n} \leqslant \alpha\|L^{-1}\|\|Lx\|_{\mathbb{R}^n} = \|Lx\|_{\mathbb{R}^n} \leqslant \|(L-K)x\|_{\mathbb{R}^n} + \|Kx\|_{\mathbb{R}^n} \\ &\leqslant \beta\|x\|_{\mathbb{R}^n} + \|Kx\|_{\mathbb{R}^n} .\end{aligned}$$

As a consequence, $(\alpha - \beta)\|x\|_{\mathbb{R}^n} \leqslant \|Kx\|_{\mathbb{R}^n}$ and this implies that $K : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one hence invertible.

2. By 1, we find that if $\|K - L\| < \dfrac{1}{\|L^{-1}\|}$, then $K \in \mathrm{GL}(n)$. Then $D\big(L, \frac{1}{\|L^{-1}\|}\big) \subseteq \mathrm{GL}(n)$ if $L \in \mathrm{GL}(n)$. Therefore, $\mathrm{GL}(n)$ is open.

3. Let $L \in \mathrm{GL}(n)$ and $\varepsilon > 0$ be given. Choose $\delta = \min\Big\{\dfrac{1}{2\|L^{-1}\|}, \dfrac{\varepsilon}{2\|L^{-1}\|^2}\Big\}$. If $\|K - L\| < \delta$, then $K \in \mathrm{GL}(n)$. Since $L^{-1} - K^{-1} = K^{-1}(K - L)L^{-1}$, we find that if $\|K - L\| < \delta$,

$$\|K^{-1}\| - \|L^{-1}\| \leqslant \|K^{-1} - L^{-1}\| \leqslant \|K^{-1}\|\|K - L\|\|L^{-1}\| < \frac{1}{2}\|K^{-1}\|$$

which implies that $\|K^{-1}\| < 2\|L^{-1}\|$. Therefore, if $\|K - L\| < \delta$,

$$\|L^{-1} - K^{-1}\| \leqslant \|K^{-1}\|\|K - L\|\|L^{-1}\| < 2\|L^{-1}\|^2\delta < \varepsilon .$$ □

Remark 6.9. There is another way to see that $\mathrm{GL}(n)$ is open in $\mathscr{B}(\mathbb{R}^n,\mathbb{R}^n)$. Let $\mathcal{M}(n)$ be the collection of $n\times n$ real matrices, and $\|\cdot\|_2$ be the matrix norm introduced in Example 1.138. Also define $\|\cdot\|:\mathcal{M}(n)\to\mathbb{R}$ by

$$\|A\| = \max\left\{|a_{ij}|\,\middle|\,A=[a_{ij}]1\leqslant i,j\leqslant n\right\}.$$

Then $\|\cdot\|$ is also a norm on $\mathcal{M}(n)$. Since $\mathcal{M}(n)$ is finite dimensional (in fact, $\dim\mathcal{M}(n)=n^2$), by Example 4.24 $\|\cdot\|$ and $\|\cdot\|_2$ are equivalent norms on $\mathcal{M}(n)$; that is, there exists $C,c>0$ such that

$$c\|A\|\leqslant\|A\|_2\leqslant C\|A\| \qquad \forall\, A\in\mathcal{M}(n).$$

Let $\{A_k\}_{k=1}^{\infty}\subseteq\mathcal{M}(n)$ be a sequence of $n\times n$ real matrices. The equivalence between $\|\cdot\|$ and $\|\cdot\|_2$ implies that $A_k\to A$ in $\mathcal{M}(n)$ if and only if each entry of A_k converges to corresponding entry of A. Therefore, the determinant function is continuous on $\mathcal{M}(n)$. In other words,

$$\lim_{A_k\to A}\det(A_k)=\det(A) \qquad \forall\, A\in\mathcal{M}(n).$$

Since $\mathrm{GL}(n)$ can be viewed as the collection of $n\times n$ matrices with non-zero determinant; that is,

$$\mathrm{GL}(n)=\left\{A\in\mathcal{M}(n)\,\middle|\,\det(A)\neq 0\right\},$$

by the continuity of the determinant function and Theorem 4.11, we conclude that $\mathrm{GL}(n)$ is open in $\mathcal{M}(n)$.

6.1.1 The matrix representation of linear maps between finite dimensional normed spaces

Let $(X,\|\cdot\|_X)$ and $(Y,\|\cdot\|_Y)$ be two finite dimensional normed spaces. Suppose that $\mathcal{B}=\{\mathrm{e}_k\}_{k=1}^n$ and $\widetilde{\mathcal{B}}=\{\widetilde{\mathrm{e}}_k\}_{k=1}^m$ are basis of X and Y, respectively. Then every $x\in X$, and $y\in Y$, there exists unique vectors $(c_1,\cdots,c_n)\in\mathbb{R}^n$ and $(d_1,\cdots,d_m)\in\mathbb{R}^m$ such that

$$x=c_1\mathrm{e}_1+\cdots+c_n\mathrm{e}_n \quad\text{and}\quad y=d_1\widetilde{\mathrm{e}}_1+\cdots+d_m\widetilde{\mathrm{e}}_m.$$

We write $[x]_{\mathcal{B}}$ for the column vector $[c_1,\cdots,c_n]^{\mathrm{T}}$ and $[y]_{\widetilde{\mathcal{B}}}$ for the column vector $[d_1,\cdots,d_m]^{\mathrm{T}}$. Then for each $L\in\mathscr{L}(X,Y)$, the matrix representation of L with respect to basis $\mathcal{B}$ and $\widetilde{\mathcal{B}}$, denoted by $[L]_{\mathcal{B},\widetilde{\mathcal{B}}}$, is the matrix $\Big[[L\mathrm{e}_1]_{\widetilde{\mathcal{B}}}\,\vdots\,[L\mathrm{e}_2]_{\widetilde{\mathcal{B}}}\,\vdots\,\cdots\,\vdots\,[L\mathrm{e}_n]_{\widetilde{\mathcal{B}}}\Big]$. The matrix $[L]_{\mathcal{B},\widetilde{\mathcal{B}}}$ has the property that

$$[Lx]_{\widetilde{\mathcal{B}}}=[L]_{\mathcal{B},\widetilde{\mathcal{B}}}[x]_{\mathcal{B}}.$$

6.2 Definition of Derivatives and the Jacobian Matrices

Definition 6.10. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces. A map $f : A \subseteq X \to Y$ is said to be ***differentiable*** at $x_0 \in A$ if there is a bounded linear map in $\mathscr{B}(X, Y)$, denoted by $(Df)(x_0)$ and called the ***derivative*** of f at x_0, such that

$$\lim_{\substack{x \to x_0 \\ x \in A}} \frac{\big\|f(x) - f(x_0) - (Df)(x_0)(x - x_0)\big\|_Y}{\|x - x_0\|_X} = 0\,,$$

where $(Df)(x_0)(x - x_0)$ denotes the value of the linear map $(Df)(x_0)$ applied to the vector $x - x_0 \in X$ (so $(Df)(x_0)(x - x_0) \in Y$). In other words, f is differentiable at $x_0 \in A$ if there exists $L \in \mathscr{B}(X, Y)$ such that

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 \ni \|f(x) - f(x_0) - L(x - x_0)\|_Y \leqslant \varepsilon \|x - x_0\|_X \text{ whenever } x \in D(x_0, \delta) \cap A\,.$$

If f is differentiable at each point of A, we say that f is differentiable on A.

Remark 6.11. Suppose that $f : A \to Y$ is differentiable on A, then Df itself is a map from A to $\mathscr{B}(X, Y)$. For each $x \in A$, $Df(x)$ is a linear map, but Df in general is not linear in x.

Example 6.12. Let $f : (0, \infty) \to \mathbb{R}$ be given by $f(x) = \dfrac{1}{x}$. Then f is differentiable at any $x_0 \in (0, \infty)$ since $(Df)(x_0) : \mathbb{R} \to \mathbb{R}$ is the linear map given by

$$(Df)(x_0)(x) = -\frac{1}{{x_0}^2} \cdot x\,.$$

To see this, we observe that

$$\lim_{x \to x_0} \frac{\left|\frac{1}{x} - \frac{1}{x_0} - \frac{-1}{{x_0}^2}(x - x_0)\right|}{|x - x_0|} = \lim_{x \to x_0} \frac{\left|\frac{{x_0}^2 - x x_0 + x^2 - x x_0}{x {x_0}^2}\right|}{|x - x_0|} = \lim_{x \to x_0} \frac{{x_0}^2 - 2 x x_0 + x^2}{x {x_0}^2 |x - x_0|}$$
$$= \lim_{x \to x_0} \frac{|x - x_0|}{x {x_0}^2} = 0\,.$$

Remark 6.13. Let $f : (a, b) \to \mathbb{R}$ be "differentiable" at $x_0 \in (a, b)$ in the sense of Definition 4.55. The "derivative" $f'(x_0)$ and the derivative $(Df)(x_0)$ is related by $(Df)(x_0)(h) = f'(x_0)h$ since

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

Example 6.14. View $(\mathbb{C}, |\cdot|)$ as a normed vector space (over field $\mathbb{C}$). Then $f : \mathbb{C} \to \mathbb{R}$ given by $f(z) = |z|^2$ is differentiable at $z_0 = 0$ and $(Df)(0) = 0$ since

$$\lim_{z\to 0} \frac{\big||z|^2 - |0|^2 - 0\cdot(z-0)\big|}{|z|} = \lim_{z\to 0} |z| = 0\,.$$

However, f is not differentiable at any $z_0 \neq 0$. In fact, in Exercise Problem 6.3 one is asked to show that if $f : \mathbb{C} \to \mathbb{R}$ is differentiable at z_0, then $(Df)(z_0) = 0$. Therefore, if f is differentiable at $z_0 \neq 0$, then

$$\begin{aligned}\frac{\big||z|^2 - |z_0|^2 - 0\cdot(z-z_0)\big|}{|z-z_0|} &= \Big|\frac{z\cdot\bar{z} - z_0\cdot\bar{z_0}}{z-z_0}\Big| = \Big|\frac{(z-z_0)\cdot\bar{z_0} + z\cdot\overline{(z-z_0)}}{z-z_0}\Big| \\ &= \Big|\bar{z_0} + \frac{z\cdot\overline{(z-z_0)}}{z-z_0}\Big| = \Big|\bar{z_0} + \overline{z-z_0} + \frac{z_0\cdot\overline{(z-z_0)}}{z-z_0}\Big|\end{aligned}$$

and the limit of the right-hand side as z approaches z_0 does not exist since $\displaystyle\lim_{z\to z_0} \frac{z_0\cdot\overline{(z-z_0)}}{z-z_0}$ does not exist (by the fact that the limit as z approaches z_0 from the horizontal and vertical directions are different).

On the other hand, the function $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(x,y) \equiv f(x+iy) = x^2 + y^2$ is differentiable everywhere and $(Dg)(a,b)\boldsymbol{v} = 2av_1 + 2bv_2$ for all $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$. To see this,

$$\frac{(a+h)^2 + (b+k)^2 - (a^2+b^2) - (2ah+2bk)}{\sqrt{h^2+k^2}} = \frac{h^2+k^2}{\sqrt{h^2+k^2}} \to 0 \quad \text{as} \quad (h,k) \to (0,0)$$

which implies that $\displaystyle\lim_{(x,y)\to(a,b)} \frac{\big\|(x^2+y^2) - (a^2+b^2) - (Dg)(a,b)(x-a,y-b)\big\|_{\mathbb{R}^2}}{\|(x-a,y-b)\|_{\mathbb{R}^2}} = 0.$

Example 6.15. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces. Then every bounded linear map $L : X \to Y$ is differentiable. In fact, $(DL)(x_0) = L$ for all $x_0 \in X$ since

$$\lim_{x\to x_0} \frac{\|Lx - Lx_0 - L(x-x_0)\|_Y}{\|x-x_0\|_X} = 0\,.$$

Example 6.16. Let $f : \mathrm{GL}(n) \to \mathrm{GL}(n)$ be given by $f(L) = L^{-1}$, where $\mathrm{GL}(n)$ is defined in Theorem 6.8. Then f is differentiable at any "point" $L \in \mathrm{GL}(n)$ with derivative $(Df)(K) \in \mathscr{B}(\mathrm{GL}(n), \mathrm{GL}(n))$ given by $(Df)(L)(K) = -L^{-1}KL^{-1}$ for all $K \in \mathrm{GL}(n)$. The proof is left as an exercise.

Theorem 6.17. *Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be normed vector spaces, $\mathcal{U} \subseteq X$ be an open set, and $f : \mathcal{U} \to Y$ be differentiable at $x_0 \in \mathcal{U}$. Then $(Df)(x_0)$ is uniquely determined by f.*

Proof. Suppose $L_1, L_2 \in \mathscr{B}(X,Y)$ are derivatives of f at x_0. Let $\varepsilon > 0$ be given and $\mathrm{e} \in X$ be a unit vector; that is, $\|\mathrm{e}\|_X = 1$. Since $\mathcal{U}$ is open, there exists $r > 0$ such that $D(x_0, r) \subseteq \mathcal{U}$. By Definition 6.10, there exists $0 < \delta < r$ such that

$$\frac{\|f(x) - f(x_0) - L_1(x - x_0)\|_Y}{\|x - x_0\|_X} < \frac{\varepsilon}{2} \quad \text{and} \quad \frac{\|f(x) - f(x_0) - L_2(x - x_0)\|_Y}{\|x - x_0\|_X} < \frac{\varepsilon}{2}$$

if $0 < \|x - x_0\|_X < \delta$. Letting $x = x_0 + \lambda\mathrm{e}$ with $0 < |\lambda| < \delta$, we have

$$\begin{aligned}\|L_1\mathrm{e} - L_2\mathrm{e}\|_Y &= \frac{1}{|\lambda|}\|L_1(x - x_0) - L_2(x - x_0)\|_Y \\ &\leqslant \frac{1}{|\lambda|}\big(\big\|f(x) - f(x_0) - L_1(x - x_0)\big\|_Y + \big\|f(x) - f(x_0) - L_2(x - x_2)\big\|_Y\big) \\ &= \frac{\big\|f(x) - f(x_0) - L_1(x - x_0)\big\|_Y}{\|x - x_0\|_X} + \frac{\big\|f(x) - f(x_0) - L_2(x - x_0)\big\|_Y}{\|x - x_0\|_X} \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.\end{aligned}$$

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

Example 6.18. $(Df)(x_0)$ 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 6.10 since

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

Remark 6.19. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be an open set and suppose that $f : \mathcal{U} \to \mathbb{R}^m$ is differentiable on $\mathcal{U}$. Then $Df : \mathcal{U} \to \mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)$. Treating Df as a map from $\mathcal{U}$ to the normed space $\big(\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m), \|\cdot\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)}\big)$, and suppose that Df is also differentiable on $\mathcal{U}$. Then the derivative of Df, denoted by $D^2 f$, is a map from $\mathcal{U}$ to $\mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n,\mathbb{R}^m))$. In other words, for each $a \in \mathcal{U}$, $(D^2 f)(a) \in \mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n,\mathbb{R}^m))$ satisfying

$$\lim_{x\to a} \frac{\big\|(Df)(x) - (Df)(a) - (D^2 f)(a)(x-a)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)}}{\|x-a\|_{\mathbb{R}^n}} = 0\,,$$

here $(D^2 f)(a)$ is bounded linear map from $\mathbb{R}^n$ to $\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)$; thus $(D^2 f)(a)(x-a) \in \mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)$.

Definition 6.20. Let $\{e_k\}_{k=1}^n$ be the standard basis of $\mathbb{R}^n$, $\mathcal{U} \subseteq \mathbb{R}^n$ 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 in the direction e_j, denoted by $\dfrac{\partial f}{\partial x_j}(a)$, is the limit

$$\lim_{h \to 0} \frac{f(a + he_j) - f(a)}{h}$$

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 6.21. *Suppose $\mathcal{U} \subseteq \mathbb{R}^n$ is an open set and $f : \mathcal{U} \to \mathbb{R}^m$ 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 map $Df(a)$ with respect to the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$ is given by*

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

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

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

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

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

where $(Le_j)_i$ denotes the i-th component of Le_j in the standard basis. As a consequence, for each $i = 1, \cdots, m$,

$$\lim_{h \to 0} \frac{f_i(a + he_j) - f_i(a)}{h} = (Le_j)_i \text{ exists}$$

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

Definition 6.22. 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). If $n = m$, the determinant of $(Jf)(x)$ is called the ***Jacobian*** of f at x.

Remark 6.23. 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, $\big[(Df)(x)\big] = (Jf)(x)$.

Example 6.24. 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

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

Example 6.25. 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 $\big[(Df)(0,0)\big] = \big[0 \quad 0\big]$. However,

$$\Big| f(x,y) - f(0,0) - \big[0 \quad 0\big] \begin{bmatrix} x \\ y \end{bmatrix} \Big| = \frac{|xy|}{x^2+y^2} = \frac{|xy|}{(x^2+y^2)^{\frac{3}{2}}} \sqrt{x^2+y^2}\,;$$

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 6.26. 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 $\big[(Df)(0,0)\big] = \big[1 \quad 1\big]$. However,

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

thus if $xy \neq 0$,

$$\big|f(x,y) - (x+y)\big| = |1-x-y| \not\to 0 \text{ as } (x,y) \to (0,0), xy \neq 0.$$

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

6.3 Conditions for Differentiability

Proposition 6.27. *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 Theorem 6.7, 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 Theorem 6.7, 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 $\|x-a\|_{\mathbb{R}^n} < \delta = \min\big\{\delta_1, \cdots, \delta_m\big\}$. □

Theorem 6.28. *Let* $\mathcal{U} \subseteq \mathbb{R}^n$ *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* $(n-1)$ *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 $\dfrac{\partial f}{\partial x_1}, \dfrac{\partial f}{\partial x_2}, \cdots, \dfrac{\partial f}{\partial x_{n-1}}$ are continuous at a. Let $\{\mathrm{e}_j\}_{j=1}^n$ be the standard basis of $\mathbb{R}^n$, and $\varepsilon > 0$ be given. Since $\dfrac{\partial f}{\partial x_i}$ is continuous at a for $i = 1, \cdots, n-1$,

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

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 $\|x-a\|_{\mathbb{R}^n} = \|k\|_{\mathbb{R}^n} < \delta$,

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

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,

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

which implies that f is differentiable at a. ▫

Remark 6.29. 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 6.25, 6.26, 6.41 are discontinuous at $(0,0)$, and these functions are not differentiable at $(0,0)$.

Example 6.30. Let $\mathcal{U} = \mathbb{R}^2\backslash\{(x,0)\in\mathbb{R}^2\,|\,x\geqslant 0\}$, 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 6.31. 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 $Df:\mathcal{U}\to\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)$ is continuous on $\mathcal{U}$. The collection of all continuously differentiable mappings 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 derivative is continuous and bounded is denoted by $\mathscr{C}^1_b(\mathcal{U};\mathbb{R}^m)$. In other words,

$$\mathscr{C}^1(\mathcal{U};\mathbb{R}^m) = \big\{f:\mathcal{U}\to\mathbb{R}^m \text{ is differentiable on } \mathcal{U}\,\big|\,Df:\mathcal{U}\to\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)\text{ is continuous}\big\}$$

and

$$\mathscr{C}^1_b(\mathcal{U};\mathbb{R}^m) = \Big\{ f \in \mathscr{C}^1(\mathcal{U};\mathbb{R}^m) \,\Big|\, \sup_{x\in\mathcal{U}} |f(x)| + \sup_{x\in\mathcal{U}} \|Df(x)\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)} < \infty \Big\}\,.$$

Corollary 6.32. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}^m$. Then $f \in \mathscr{C}^1(\mathcal{U};\mathbb{R}^m)$ if and only if the partial derivatives $\dfrac{\partial f_i}{\partial x_j}$ exist and are continuous on $\mathcal{U}$ for $i = 1, \cdots, m$ and $j = 1, \cdots, n$.*

Proof. Note that for any matrix $A = [a_{ij}]_{m\times n}$, $\|A\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)} \leqslant \sum\limits_{i,j} |a_{ij}| \leqslant nm\|A\|$; thus

$$\begin{aligned}
\big\|(Df)(x) - (Df)(x_0)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)} &\leqslant \sum_{i=1}^m \sum_{j=1}^n \Big|\frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(x_0)\Big| \\
&\leqslant nm\big\|(Df)(x) - (Df)(x_0)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)}\,.
\end{aligned}$$

Therefore, the continuity of Df is equivalent to the continuity of the partial derivatives $\dfrac{\partial f_i}{\partial x_j}$ for all i, j. The corollary is then concluded by Proposition 6.27 and Theorem 6.28. □

Example 6.33. 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

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

Then f is differentiable at $x = 0$ since the limit

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

exists. Therefore,

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

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

Proposition 6.34. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open. Given $f \in \mathscr{C}^1_b(\mathcal{U};\mathbb{R}^m)$, define*

$$\|f\|_{\mathscr{C}^1_b(\mathcal{U};\mathbb{R}^m)} = \sup_{x\in\mathcal{U}} \Big[|f(x)| + \sum_{i=1}^m \sum_{j=1}^n \Big|\frac{\partial f_i}{\partial x_j}(x)\Big| \Big]\,.$$

Then $\big(\mathscr{C}^1_b(\mathcal{U};\mathbb{R}^m), \|\cdot\|_{\mathscr{C}^1_b(\mathcal{U};\mathbb{R}^m)}\big)$ is a Banach space.

Proof. Left as an exercise. □

Theorem 6.35. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $K \subseteq \mathcal{U}$ be compact, and $f : \mathcal{U} \to \mathbb{R}$ be of class $\mathscr{C}^1$. Then for each $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\big|f(y) - f(x) - (Df)(x)(y-x)\big| \leqslant \varepsilon \|y-x\|_{\mathbb{R}^n} \quad \text{if } \|y-x\|_{\mathbb{R}^n} < \delta \text{ and } x, y \in K\,.$$

Proof. Define $g : \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ by

$$g(x,y) = \begin{cases} \dfrac{\big|f(y) - f(x) - (Df)(x)(y-x)\big|}{\|y-x\|_{\mathbb{R}^n}} & \text{if } y \neq x\,, \\ 0 & \text{if } y = x\,. \end{cases}$$

Since f is of class $\mathscr{C}^1$, g is continuous on $\mathcal{U} \times \mathcal{U}$. In fact, it is clear that g is continuous at (x,y) if $x \neq y$, while the mean value theorem implies that $f(w) - f(z) = (Df)(\xi)(w-z)$ for some ξ on the line segment joining w and z; thus

$$\begin{aligned} &\limsup_{\substack{(z,w)\to(x,x)\\ z\neq w}} \frac{\big|f(w) - f(z) - (Df)(z)(w-z)\big|}{\|w-z\|_{\mathbb{R}^n}} \\ &\quad = \limsup_{\substack{(z,w)\to(x,x)\\ z\neq w}} \frac{\big|\big((Df)(\xi) - (Df)(z)\big)(w-z)\big|}{\|w-z\|_{\mathbb{R}^n}} \leqslant \limsup_{\substack{(z,w)\to(x,x)\\ z\neq w}} \big\|(Df)(\xi) - (Df)(z)\big\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R})} = 0\,. \end{aligned}$$

Now by the compactness of $K \times K$, for each given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|g(z,w) - g(x,y)| < \varepsilon \qquad \text{if } \|(z,w) - (x,y)\|_{\mathbb{R}^{2n}} < \delta \text{ and } x,y,z,w \in K\,.$$

In particular, with $(z,w) = (x,x)$ we find that $|g(x,y)| < \varepsilon$ if $\|x-y\|_{\mathbb{R}^n} < \delta$; thus

$$\frac{\big|f(y) - f(x) - (Df)(x)(y-x)\big|}{\|y-x\|_{\mathbb{R}^n}} < \varepsilon \qquad \text{if } 0 < \|x-y\|_{\mathbb{R}^n} < \delta\,,\ x,y \in K\,.$$ □

Corollary 6.36. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $K \subseteq \mathcal{U}$ be compact, and $f : \mathcal{U} \to \mathbb{R}^m$ be of class $\mathscr{C}^1$. Then for each $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\big\|f(y) - f(x) - (Df)(x)(y-x)\big\|_{\mathbb{R}^m} \leqslant \varepsilon \|y-x\|_{\mathbb{R}^n} \quad \text{if } \|y-x\|_{\mathbb{R}^n} < \delta \text{ and } x, y \in K\,.$$

6.4 Properties of Differentiable Functions

6.4.1 Continuity of differentiable maps

Theorem 6.37. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces, $\mathcal{U} \subseteq X$ be open, and $f : \mathcal{U} \to Y$ 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 $L \in \mathscr{B}(X, Y)$ such that

$$\exists\, \delta_1 > 0 \ni \left\|f(x) - f(x_0) - L(x - x_0)\right\|_Y \leqslant \|x - x_0\|_X \quad \forall\, x \in D(x_0, \delta_1)\,.$$

As a consequence,

$$\left\|f(x) - f(x_0)\right\|_Y \leqslant \big(\|L\| + 1\big)\|x - x_0\|_X \quad \forall\, x \in D(x_0, \delta_1)\,. \tag{6.4.1}$$

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

$$\left\|f(x) - f(x_0)\right\|_Y \leqslant \frac{\varepsilon}{2} < \varepsilon\,.$$ □

Remark 6.38. 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 (6.4.1)).

Example 6.39. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given in Example 6.25. 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 6.40. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given in Example 6.26. Then f is not continuous at $(0, 0)$; thus not differentiable at $(0, 0)$.

Example 6.41. 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

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

6.4.2 The product rules

Proposition 6.42. *Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ be normed vector spaces, $\mathcal{U} \subseteq X$ be open, and $f : \mathcal{U} \to Y$ and $g : \mathcal{U} \to \mathbb{F}$ be differentiable at $x_0 \in \mathcal{U}$, where $\mathbb{F}$ is the scalar field associated with the vector space Y. Then $gf : \mathcal{U} \to Y$ is differentiable at x_0, and*

$$D(gf)(x_0)(v) = g(x_0)(Df)(x_0)(v) + (Dg)(x_0)(v)f(x_0)\,. \tag{6.4.2}$$

Moreover, if $g(x_0) \neq 0$, then $\dfrac{f}{g} : \mathcal{U} \to Y$ is also differentiable at x_0, and $D(\dfrac{f}{g})(x_0) : X \to Y$ is given by

$$D\big(\frac{f}{g}\big)(x_0)(v) = \frac{g(x_0)\big((Df)(x_0)(v)\big) - (Dg)(x_0)(v)f(x_0)}{g^2(x_0)}\,. \tag{6.4.3}$$

Proof. We only prove (6.4.2), and (6.4.3) is left as an exercise.

Define $Av = g(x_0)(Df)(x_0)(v) + (Dg)(x_0)(v)f(x_0)$. Then $A \in \mathscr{B}(X, Y)$. Moreover,

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

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

$$\big|(Dg)(x_0)(x - x_0)\big| \leqslant \big\|(Dg)(x_0)\big\|_{\mathscr{B}(X,\mathbb{F})}\|x - x_0\|_X$$

and the continuity of f at x_0 (due to Theorem 6.37), we find that

$$\lim_{x\to x_0}\Big|\frac{\big|(Dg)(x_0)(x - x_0)\big|}{\|x - x_0\|_X}\big\|f(x) - f(x_0)\big\|_Y\Big| \leqslant \lim_{x\to x_0}\big\|(Dg)(x_0)\big\|_{\mathscr{B}(X,\mathbb{F})}\big\|f(x) - f(x_0)\big\|_Y = 0\,.$$

As a consequence,

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

which implies that gf is differentiable at x_0 with derivative $D(gf)(x_0)$ given by (6.4.2). □

6.4.3 The chain rule

Theorem 6.43. *Let* $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$, $(Z, \|\cdot\|_Z)$ *be normed vector spaces,* $\mathcal{U} \subseteq X$ *and* $\mathcal{V} \subseteq Y$ *be open sets. Suppose that* $f : \mathcal{U} \to Y$ *is differentiable at* $x_0 \in \mathcal{U}$, $f(\mathcal{U}) \subseteq \mathcal{V}$, *and* $g : \mathcal{V} \to Z$ *is differentiable at* $f(x_0)$. *Then the map* $F = g \circ f : \mathcal{U} \to Z$ *defined by*

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

is differentiable at x_0, *and*

$$(DF)(x_0)(h) = (Dg)\big(f(x_0)\big)\big((Df)(x_0)(h)\big) \qquad \forall\, h \in X\,.$$

In particular, if $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ *and* $Z = \mathbb{R}^\ell$, *then*

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

Proof. To simplify the notation, let $y_0 = f(x_0)$, $A = (Df)(x_0) \in \mathscr{B}(X,Y)$, and $B = (Dg)(y_0) \in \mathscr{B}(Y,Z)$. 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 $\|x - x_0\|_X < \delta_1$ and $\|y - y_0\|_Y < \delta_2$, we have

$$\begin{aligned}\|f(x) - f(x_0) - A(x - x_0)\|_Y &\leqslant \min\Big\{1, \frac{\varepsilon}{2(\|B\|+1)}\Big\}\|x - x_0\|_X\,,\\ \|g(y) - g(y_0) - B(y - y_0)\|_Z &\leqslant \frac{\varepsilon}{2(\|A\|+1)}\|y - y_0\|_Y\,.\end{aligned}$$

Define

$$\begin{aligned}u(h) &= f(x_0 + h) - f(x_0) - Ah \qquad \forall\, \|h\|_X < \delta_1\,,\\ v(k) &= g(y_0 + k) - g(y_0) - Bk \qquad \forall\, \|k\|_Y < \delta_2\,.\end{aligned}$$

Then if $\|h\|_X < \delta_1$ and $\|k\|_Y < \delta_2$,

$$\|u(h)\|_Y \leqslant \|h\|_X\,, \quad \|u(h)\|_Y \leqslant \frac{\varepsilon}{2(\|B\|+1)}\|h\|_X \quad \text{and} \quad \|v(k)\|_Z \leqslant \frac{\varepsilon}{2(\|A\|+1)}\|k\|_Y\,.$$

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

$$\|k\|_Y < \delta_2 \quad \text{whenever} \quad \|h\|_X < \delta_3\,.$$

Since

$$\begin{aligned}F(x_0 + h) - F(x_0) &= g(y_0 + k) - g(y_0) = Bk + v(k) = B\big(Ah + u(h)\big) + v(k)\\ &= BAh + Bu(h) + v(k)\,,\end{aligned}$$

we conclude that if $\|h\|_X < \delta = \min\{\delta_1, \delta_3\}$,

$$\begin{aligned}\|F(x_0+h) - F(x_0) - BAh\|_Z &\leqslant \|Bu(h)\|_Z + \|v(k)\|_Z \leqslant \|B\|\|u(h)\|_Y + \frac{\varepsilon}{2(\|A\|+1)}\|k\|_Y \\ &\leqslant \frac{\varepsilon}{2}\|h\|_X + \frac{\varepsilon}{2(\|A\|+1)}\big(\|A\|\|h\|_X + \|u(h)\|_Y\big) \leqslant \frac{\varepsilon}{2}\|h\|_X + \frac{\varepsilon}{2}\|h\|_X = \varepsilon\|h\|_X\end{aligned}$$

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

Example 6.44. 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

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

which is commonly seen in Calculus textbook.

Example 6.45. 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 6.46. 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 6.47. 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

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

Way II: Use the chain rule:

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

Example 6.48. 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,

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

which implies that

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

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

6.4.4 The mean value theorem

Theorem 6.49. *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 6.43, 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 (Corollary 4.65) 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 (Corollary 4.65) 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 $g(1) = f(y)$,

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

which concludes the theorem. □

Example 6.50. 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 6.51. 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 6.52. Let f be given in Example 6.30, 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 6.53. Suppose that $A \subseteq \mathbb{R}^n$ is an open convex set, and $f : A \to \mathbb{R}^m$ is differentiable and $Df(x) = 0$ for all $x \in A$. Then f is a constant; that is, for some $\alpha \in \mathbb{R}^m$ we have $f(x) = \alpha$ for all $x \in A$.

Reason: Since A is convex, then the Mean Value Theorem can be applied to any $x, y \in A$ 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 $x, y \in A$. Let $\alpha = f(x) \in \mathbb{R}^m$, then we reach the conclusion.

Example 6.54. Let $f : [0, \infty) \to \mathbb{R}$ be continuous and be differentiable on $(0, \infty)$. Suppose that $f(0) = 0$ and $f'(x)$ is non-decreasing (that is, if $x < y$, then $f'(x) \leqslant f'(y)$). Show that $g : (0, \infty) \to \mathbb{R}$, $g(x) = \dfrac{f(x)}{x}$ is also non-decreasing.

Proof: It suffices to prove $g'(x) \geqslant 0$. Since f is differentiable on $(0, \infty)$, then g is differentiable on $(0, \infty)$, and $g'(x) = \dfrac{xf'(x) - f(x)}{x^2}$. Hence

$$g'(x) \geqslant 0 \Leftrightarrow xf'(x) \geqslant f(x).$$

Let $x > 0$ be fixed. Applying the Mean Value Theorem to f we find that

$$\exists\, c \in (0, x) \ni f(x) = f(x) - f(0) = f'(c)(x - 0) \leqslant xf'(x).$$

6.5 Directional Derivatives and Gradient Vectors

Definition 6.55. 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 6.56. 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.

Theorem 6.57. *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. Let $\varepsilon > 0$ be given. Since f is differentiable at x_0, there exists $\delta > 0$ 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 $\mathbb{R}^n$ and $0 < |t| < \delta$, then

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

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

Remark 6.58. When $v \in \mathbb{R}^n$ but $0 < \|v\|_{\mathbb{R}^n} \neq 1$, we let $\mathrm{v} = \dfrac{v}{\|v\|_{\mathbb{R}^n}}$. Then the direction derivatives of a function $f : \mathcal{U} \subseteq \mathbb{R}^n \to \mathbb{R}$ 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 $s = \dfrac{t}{\|v\|_{\mathbb{R}^n}}$. Then

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

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

Example 6.59. 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 6.41, 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 $(Jf)(0) = \Big[\frac{\partial f}{\partial x}(0,0)\ \ \frac{\partial f}{\partial y}(0,0)\Big] = \big[1\ 0\big]$ is the Jacobian matrix of f at $(0,0)$.

Example 6.60. 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 6.61. 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 6.62. 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 grad$f$ or ∇f.

Let $\mathcal{U} \subseteq \mathbb{R}^n$ 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 $1 \leqslant k \leqslant n$. W.L.O.G., we can assume that $\dfrac{\partial f}{\partial x_n}(a) \neq 0$. By the implicit function theorem, there exists an open neighborhood $\mathcal{V} \subseteq \mathbb{R}^{n-1}$ of $(a_1, \cdots, a_{n-1})$ and an open neighborhood $\mathcal{W} \subseteq \mathbb{R}$ of a_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 $x_n = \varphi(x_1, \cdots, x_{n-1})$; that is,

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

Moreover,

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

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

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

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

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

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

Proposition 6.63. *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 6.64. 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 6.65. 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

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

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 6.66. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then $\pm\dfrac{\nabla f}{\|\nabla f\|_{\mathbb{R}^n}}$ is the direction in which the function f increases/decreases most rapidly*（最速上升／下降方向）.

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 6.67. 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 $(0,0,1)$.

6.6 Higher Order Derivatives of Functions

Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}^m$ is differentiable. By Proposition 6.5, the space $\big(\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m), \|\cdot\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)}\big)$ is a normed space (in fact, it is a Banach space), so it is legitimate to ask if $Df : \mathcal{U} \to \mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)$ is differentiable or not. If Df is differentiable at x_0, we call f twice differentiable at x_0, and denote the twice derivative of f at x_0 as $(D^2 f)(x_0)$. If Df is differentiable on $\mathcal{U}$, then $D^2 f : \mathcal{U} \to \mathscr{B}\big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)\big)$. Similar, we can talk about three times differentiability of a function if it is twice differentiable. In general, we have the following

Definition 6.68. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces, and $\mathcal{U} \subseteq X$ be open. A function $f : \mathcal{U} \to Y$ is said to be ***twice differentiable*** at $a \in \mathcal{U}$ if

1. f is (once) differentiable in a neighborhood of a;

2. there exists $L_2 \in \mathscr{B}\big(X, \mathscr{B}(X, Y)\big)$, usually denoted by $(D^2 f)(a)$ and called the ***second derivative*** of f at a, such that

$$\lim_{x \to a} \frac{\big\|(Df)(x) - (Df)(a) - L_2(x-a)\big\|_{\mathscr{B}(X,Y)}}{\|x-a\|_X} = 0\,.$$

For any two vectors $u, v \in X$, $(D^2 f)(a)(v) \in \mathscr{B}(X, Y)$ and $(D^2 f)(a)(v)(u) \in Y$. The vector $(D^2 f)(a)(v)(u)$ is usually denoted by $(D^2 f)(a)(u, v)$.

In general, a function f is said to be k***-times differentiable*** at $a \in \mathcal{U}$ if

1. f is $(k-1)$-times differentiable in a neighborhood of a;

2. there exists $L_k \in \mathscr{B}(\underbrace{X, \mathscr{B}(X, \cdots, \mathscr{B}(X}_{k \text{ copies of "}X\text{"}}, Y\underbrace{)\cdots))}_{k \text{ copies of ")"}}$, usually denoted by $(D^k f)(a)$ and called the k***-th derivative*** of f at a, such that

$$\lim_{x \to a} \frac{\big\|(D^{k-1} f)(x) - (D^{k-1} f)(a) - L_k(x-a)\big\|_{\mathscr{B}(X, \mathscr{B}(X, \cdots, \mathscr{B}(X,Y)\cdots))}}{\|x-a\|_X} = 0\,.$$

For any k vectors $u^{(1)}, \cdots u^{(k)} \in X$, the vector $(D^k f)(a)(u^{(1)}, \cdots, u^{(k)})$ is defined as the vector

$$(D^k f)(a)(u^{(k)})(u^{(k-1)}) \cdots (u^{(1)})\,.$$

Example 6.69. Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two normed spaces, and $f(x) = Lx$ for some $L \in \mathscr{B}(X, Y)$. From Example 6.15, $(Df)(x_0) = L$ for all $x_0 \in X$; thus $(D^2 f)(x_0) = 0$ since $Df : \mathcal{U} \in \mathscr{B}(X, Y)$ is a "constant" map. In fact, one can also conclude from

$$\lim_{x \to x_0} \frac{\big\|(Df)(x) - (Df)(x_0) - 0(x-x_0)\big\|_{\mathscr{B}(X,Y)}}{\|x-x_0\|_X} = 0$$

that $(D^2 f)(x_0) = 0$ for all $x_0 \in X$.

Remark 6.70. We focus on what $(D^k f)(a)(u_k)(\cdots)(u_1)$ means in this remark. We first look at the case that f is twice differentiable at a. With $x = a + tv$ for $v \in X$ with $\|v\|_X = 1$ in the definition, we find that

$$\lim_{t \to 0} \frac{\big\|(Df)(a+tv) - (Df)(a) - t(D^2 f)(a)(v)\big\|_{\mathscr{B}(X,Y)}}{|t|} = 0\,.$$

Since $(Df)(a+tv)-(Df)(a)-t(D^2f)(a)(v)\in\mathscr{B}(X,Y)$, for all $u\in X$ with $\|u\|_X=1$ we have

$$\begin{aligned}&\lim_{t\to 0}\frac{\|(Df)(a+tv)(u)-(Df)(a)(u)-t(D^2f)(a)(v)(u)\|_Y}{|t|}\\&\quad=\lim_{t\to 0}\frac{\big\|\big[(Df)(a+tv)-(Df)(a)-t(D^2f)(a)(v)\big](u)\big\|_Y}{|t|}\\&\quad\leqslant\lim_{t\to 0}\frac{\|(Df)(a+tv)-(Df)(a)-t(D^2f)(a)(v)\|_{\mathscr{B}(X,Y)}}{|t|}=0\,.\end{aligned}$$

On the other hand, by the definition of the direction derivative,

$$(Df)(a+tv)(u)-(Df)(a)(u)=\lim_{s\to 0}\Big[\frac{f(a+tv+su)-f(a+tv)}{s}-\frac{f(a+su)-f(a)}{s}\Big];$$

thus the limit above implies that

$$\begin{aligned}(D^2f)(a)(v)(u)&=\lim_{t\to 0}\lim_{s\to 0}\frac{f(a+tv+su)-f(a+tv)-f(a+su)+f(a)}{st}\\&=\lim_{t\to 0}\frac{\lim\limits_{s\to 0}\frac{f(a+tv+su)-f(a+tv)}{s}-\lim\limits_{s\to 0}\frac{f(a+su)-f(a)}{s}}{t}\\&=D_v(D_uf)(a)\,.\end{aligned}$$

Therefore, $(D^2f)(a)(v)(u)$ is obtained by first differentiating f near a in the u-direction, then differentiating (Df) at a in the v-direction.

In general, $(D^kf)(a)(u_k)\cdots(u_1)$ is obtained by first differentiating f near a in the u_1-direction, then differentiating (Df) near a in the u_2-direction, and so on, and finally differentiating $(D^{k-1}f)$ at a in the u_k-direction.

Remark 6.71. Since $(D^2f)(a)\in\mathscr{B}(X,\mathscr{B}(X,Y))$, if $v_1,v_2\in X$ and $c\in\mathbb{R}$, we have $(D^2f)(a)(cv_1+v_2)=c(D^2f)(a)(v_1)+(D^2f)(a)(v_2)$ (treated as "vectors" in $\mathscr{B}(X,Y)$); thus

$$(D^2f)(a)(cv_1+v_2)(u)=c(D^2f)(a)(v_1)(u)+(D^2f)(a)(v_2)(u)\qquad\forall\,u,v_1,v_2\in X\,.$$

On the other hand, since $(D^2f)(a)(v)\in\mathscr{B}(X,Y)$,

$$(D^2f)(a)(v)(cu_1+u_2)=c(D^2f)(a)(v)(u_1)+(D^2f)(a)(v)(u_2)\qquad\forall\,u_1,u_2,v\in X\,.$$

Therefore, $(D^2f)(a)(v)(u)$ is linear in both u and v variables. A map with such kind of property is called a ***bilinear*** map (meaning 2-linear). In particular, $(D^2f)(a):X\times X\to Y$ is a bilinear map.

In general, the vector $(D^k f)(a)(u^{(1)}, \cdots, u^{(k)})$ is linear in $u^{(1)}, \cdots, u^{(k)}$; that is,

$$\begin{aligned}&(D^k f)(a)(u^{(1)}, \cdots, u^{(i-1)}, \alpha v + \beta w, u^{(i+1)}, \cdots, u^{(k)})\\ &\quad = \alpha (D^k f)(a)(u^{(1)}, \cdots, u^{(i-1)}, v, u^{(i+1)}, \cdots, u^{(k)})\\ &\qquad + \beta (D^k f)(a)(u^{(1)}, \cdots, u^{(i-1)}, w, u^{(i+1)}, \cdots, u^{(k)})\end{aligned}$$

for all $v, w \in X$, $\alpha, \beta \in \mathbb{R}$, and $i = 1, \cdots, n$. Such kind of map which is linear in each component when the other $k-1$ components are fixed is called k***-linear***.

Consider the case that X is finite dimensional with $\dim(X) = n$, $\{e_1, e_2, \ldots, e_n\}$ is a basis of X, and $Y = \mathbb{R}$. Then $(D^2 f)(a) : X \times X \to Y$ is a bilinear form (here the term "form" means that $Y = \mathbb{R}$). A bilinear form $B : X \times X \to \mathbb{R}$ can be represented as follows: Let $a_{ij} = B(e_i, e_j) \in \mathbb{R}$ for $i, j = 1, 2, \cdots, n$. Given $x, y \in \mathbb{R}^n$, write $u = \sum_{i=1}^{n} u_i e_i$ and $v = \sum_{j=1}^{n} v_j e_j$. Then by the bilinearity of B,

$$B(u, v) = B\Big(\sum_{i=1}^{n} u_i e_i, \sum_{j=1}^{n} v_j e_j\Big) = \sum_{i,j=1}^{n} u_i v_j a_{ij} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

Therefore, if $f : \mathcal{U} \subseteq \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at a, then the bilinear form $(D^2 f)(a)$ can be represented as

$$(D^2 f)(a)(u, v) = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} (D^2 f)(e_1, e_1) & \cdots & (D^2 f)(a)(e_1, e_n) \\ \vdots & \ddots & \vdots \\ (D^2 f)(e_n, e_1) & \cdots & (D^2 f)(a)(e_n, e_n) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

The following proposition is an analogy of Proposition 6.27. The proof is similar to the one of Proposition 6.27, and is left as an exercise.

Proposition 6.72. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $x_0 \in \mathcal{U}$, and $f = (f_1, \cdots, f_m) : \mathcal{U} \to \mathbb{R}^m$. Then f is k-times differentiable at x_0 if and only if f_i is k-times differentiable at x_0 for all $i = 1, \cdots, m$.*

Due to the proposition above, when talking about the higher-order differentiability of $f : \mathcal{U} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and a point $x_0 \in \mathcal{U}$, from now on we only focus on the case $m = 1$.

Example 6.73. In this example, we focus on what the second derivative $(D^2 f)(a)$ of a function f is, or in particular, what $(D^2 f)(a)(e_i, e_j)$ (which appears in the Remark 6.71) is, if $X = \mathbb{R}^2$.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be differentiable, then

$$\big[(Df)(x,y)\big] = \big[f_x(x,y) \quad f_y(x,y)\big] = \left[\frac{\partial f}{\partial x}(x,y) \quad \frac{\partial f}{\partial y}(x,y)\right].$$

Suppose that f is twice differentiable at (a,b), and let $L_2 = (D^2 f)(a,b)$. Then

$$\lim_{(x,y)\to(a,b)} \frac{\big\|(Df)(x,y) - (Df)(a,b) - L_2\big((x-a,y-b)\big)\big\|_{\mathscr{B}(\mathbb{R}^2,\mathbb{R})}}{\sqrt{(x-a)^2+(y-b)^2}} = 0$$

or equivalently,

$$\lim_{(x,y)\to(a,b)} \frac{\big\|\,\big[f_x(x,y) \quad f_y(x,y)\big] - \big[f_x(a,b) \quad f_y(a,b)\big] - \big[L_2\big((x-a,y-b)\big)\big]\big\|_{\mathscr{B}(\mathbb{R}^2,\mathbb{R})}}{\sqrt{(x-a)^2+(y-b)^2}} = 0\,,$$

where $\big[L_2\big((x-a,y-b)\big)\big]$ denotes the matrix representation of the linear map $L_2\big((x-a,y-b)\big) \in \mathscr{B}(\mathbb{R}^2,\mathbb{R})$. In particular, we must have

$$\lim_{x\to a}\left\|\left[\frac{f_x(x,b)-f_x(a,b)}{x-a} \quad \frac{f_y(x,b)-f_y(a,b)}{x-a}\right] - \big[L_2\mathrm{e}_1\big]\right\|_{\mathscr{B}(\mathbb{R}^2,\mathbb{R})} = 0$$

and

$$\lim_{y\to b}\left\|\left[\frac{f_x(a,y)-f_x(a,b)}{y-b} \quad \frac{f_y(a,y)-f_y(a,b)}{y-b}\right] - \big[L_2\mathrm{e}_2\big]\right\|_{\mathscr{B}(\mathbb{R}^2,\mathbb{R})} = 0\,.$$

Using the notation of second partial derivatives, we find that

$$\big[L_2\mathrm{e}_1\big] = \big[f_{xx}(a,b) \quad f_{yx}(a,b)\big] \quad \text{and} \quad \big[L_2\mathrm{e}_2\big] = \big[f_{xy}(a,b) \quad f_{yy}(a,b)\big]\,,$$

where $f_{xy} = (f_x)_y = \dfrac{\partial}{\partial y}\Big(\dfrac{\partial f}{\partial x}\Big)$ and $f_{yx} = (f_y)_x = \dfrac{\partial}{\partial x}\Big(\dfrac{\partial f}{\partial y}\Big)$. Therefore, if $v = v_1\mathrm{e}_1 + v_2\mathrm{e}_2$,

$$[L_2 v] = \big[L_2(v_1\mathrm{e}_1 + v_2\mathrm{e}_2)\big] = \big[v_1 f_{xx}(a,b) + v_2 f_{xy}(a,b) \quad v_1 f_{yx}(a,b) + v_2 f_{yy}(a,b)\big]\,. \tag{6.6.1}$$

Symbolically, we can write

$$\big[L_2\big] = \Big[\,\big[f_{xx}(a,b) \quad f_{yx}(a,b)\big] \quad \big[f_{xy}(a,b) \quad f_{yy}(a,b)\big]\,\Big]$$

so that

$$\begin{aligned}\big[L_2(v_1\mathrm{e}_1 + v_2\mathrm{e}_2)\big] &= \big[L_2\big]\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \Big[\,\big[f_{xx}(a,b) \quad f_{yx}(a,b)\big] \quad \big[f_{xy}(a,b) \quad f_{yy}(a,b)\big]\,\Big]\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ &= v_1\,\big[f_{xx}(a,b) \quad f_{yx}(a,b)\big] + v_2\,\big[f_{xy}(a,b) \quad f_{yy}(a,b)\big]\,.\end{aligned}$$

For two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, what does $(D^2f)(a,b)(\boldsymbol{v})(\boldsymbol{u})$ or $(D^2f)(a,b)(\boldsymbol{u},\boldsymbol{v})$ mean? To see this, let $\boldsymbol{u} = u_1\mathrm{e}_1 + u_2\mathrm{e}_2$ and $\boldsymbol{v} = v_1\mathrm{e}_1 + v_2\mathrm{e}_2$. Then

$$\begin{aligned}\big[(D^2f)(a,b)(\boldsymbol{v})(\boldsymbol{u})\big] &= \big[(D^2f)(a,b)(\boldsymbol{v})\big]\big[\boldsymbol{u}\big] = \big[L_2(v_1\mathrm{e}_1+v_2\mathrm{e}_2)\big]\begin{bmatrix}u_1\\u_2\end{bmatrix}\\ &= v_1\begin{bmatrix}f_{xx}(a,b) & f_{yx}(a,b)\end{bmatrix}\begin{bmatrix}u_1\\u_2\end{bmatrix} + v_2\begin{bmatrix}f_{xy}(a,b) & f_{yy}(a,b)\end{bmatrix}\begin{bmatrix}u_1\\u_2\end{bmatrix}\\ &= \begin{bmatrix}v_1 & v_2\end{bmatrix}\begin{bmatrix}f_{xx}(a,b) & f_{yx}(a,b)\\ f_{xy}(a,b) & f_{yy}(a,b)\end{bmatrix}\begin{bmatrix}u_1\\u_2\end{bmatrix}.\end{aligned}$$

Therefore, $(D^2f)(a,b)(\mathrm{e}_1,\mathrm{e}_1) = f_{xx}(a,b)$, $(D^2f)(a,b)(\mathrm{e}_1,\mathrm{e}_2) = f_{xy}(a,b)$, $(D^2f)(a,b)(\mathrm{e}_2,\mathrm{e}_1) = f_{yx}(a,b)$ and $(D^2f)(a,b)(\mathrm{e}_2,\mathrm{e}_2) = f_{yy}(a,b)$.

On the other hand, we can identify $\mathscr{B}(\mathbb{R}^2;\mathbb{R})$ as $\mathbb{R}^2$ (every 1×2 matrix is a "row" vector), and treat $g \equiv \big[Df\big]^{\mathrm{T}} : \mathbb{R}^2 \to \mathbb{R}^2$ as a vector-valued function. By Theorem 6.21 $(Dg)(a,b)$ can be represented as a 2×2 matrix given by

$$\big[(Dg)(a,b)\big] = \begin{bmatrix}f_{xx}(a,b) & f_{xy}(a,b)\\ f_{yx}(a,b) & f_{yy}(a,b)\end{bmatrix}.$$

We note that the representation above means

$$\lim_{(x,y)\to(a,b)} \frac{\left\| \begin{bmatrix}f_x(x,y)\\ f_y(x,y)\end{bmatrix} - \begin{bmatrix}f_x(a,b)\\ f_y(a,b)\end{bmatrix} - \begin{bmatrix}f_{xx}(a,b) & f_{xy}(a,b)\\ f_{yx}(a,b) & f_{yy}(a,b)\end{bmatrix}\begin{bmatrix}x-a\\ y-b\end{bmatrix}\right\|_{\mathbb{R}^2}}{\sqrt{(x-a)^2+(y-b)^2}} = 0\,.$$

The equality above is equivalent to that

$$\lim_{(x,y)\to(a,b)} \frac{\left\|\big[(Df)(x,y)\big] - \big[(Df)(a,b)\big] - \begin{bmatrix}x-a & y-b\end{bmatrix}\begin{bmatrix}f_{xx}(a,b) & f_{yx}(a,b)\\ f_{xy}(a,b) & f_{yy}(a,b)\end{bmatrix}\right\|_{\mathbb{R}^2}}{\sqrt{(x-a)^2+(y-b)^2}} = 0$$

According to the equality above, $L_2 = (D^2f)(a,b)$ should be defined by

$$\big[L_2(v_1\mathrm{e}_1+v_2\mathrm{e}_2)\big] = \begin{bmatrix}v_1 & v_2\end{bmatrix}\begin{bmatrix}f_{xx}(a,b) & f_{yx}(a,b)\\ f_{xy}(a,b) & f_{yy}(a,b)\end{bmatrix} = \left(\begin{bmatrix}f_{xx}(a,b) & f_{xy}(a,b)\\ f_{yx}(a,b) & f_{yy}(a,b)\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix}\right)^{\mathrm{T}}$$

which agrees with what (6.6.1) provides.

Proposition 6.74. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$. Suppose that f is k-times differentiable at a. Then for k vectors $u^{(1)},\cdots,u^{(k)} \in \mathbb{R}^n$,*

$$\begin{aligned}(D^kf)(a)(u^{(1)},\cdots,u^{(k)}) &= \sum_{j_1,\cdots,j_k=1}^{n} \frac{\partial^k f}{\partial x_{j_k}\partial x_{j_{k-1}}\cdots\partial x_{j_1}}(a)u^{(1)}_{j_1}u^{(2)}_{j_2}\cdots u^{(k)}_{j_k}\\ &= \sum_{j_1,\cdots,j_k=1}^{n} \frac{\partial}{\partial x_{j_k}}\Big(\frac{\partial}{\partial x_{j_{k-1}}}\Big(\cdots\frac{\partial}{\partial x_{j_2}}\Big(\frac{\partial f}{\partial x_{j_1}}\Big)\cdots\Big)\Big)(a)u^{(1)}_{j_1}u^{(2)}_{j_2}\cdots u^{(k)}_{j_k}\,,\end{aligned}$$

where $u^{(i)} = (u_1^{(i)}, u_2^{(i)}, \cdots, u_n^{(i)})$ *for all* $i = 1, \cdots, k$.（上標括號中的數字指所給定的 k 個向量中的第幾個向量，下標指每一個固定向量的第幾個分量）

Proof. Let $\{e_j\}_{j=1}^n$ be the standard basis of $\mathbb{R}^n$. By Remark 6.71 (on multi-linearity), it suffices to show that

$$(D^k f)(a)(e_{j_k})(e_{j_{k-1}}) \cdots (e_{j_2})(e_{j_1}) = (D^k f)(a)(e_{j_1}, \cdots, e_{j_k}) = \frac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(a) \quad (6.6.2)$$

provided that f is k-times differentiable at a since if so, we must have

$$\begin{aligned}(D^k f)(a)(u^{(1)}, \cdots, u^{(k)}) &= (D^k f)(a)\Big(\sum_{j_1=1}^n u_{j_1}^{(1)} e_{j_1}, \cdots, \sum_{j_k=1}^n u_{j_k}^{(k)} e_{j_k}\Big) \\ &= \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_k=1}^n (D^k f)(a)(e_{j_1}, \cdots, e_{j_k}) u_{j_1}^{(1)} u_{j_2}^{(2)} \cdots u_{j_k}^{(k)} \\ &= \sum_{j_1, \cdots, j_k=1}^n \frac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(a) u_{j_1}^{(1)} u_{j_2}^{(2)} \cdots u_{j_k}^{(k)}.\end{aligned}$$

We prove the proposition by induction. Note that the case $k = 1$ is true because of Theorem 6.21. Next we assume that (6.6.2) holds true for $k = \ell$ if f is $(\ell-1)$-times differentiable in a neighborhood of a and f is ℓ-times differentiable at a. Now we show that (6.6.2) also holds true for $k = \ell+1$ if f is ℓ-times differentiable in a neighborhood of a, and f is $(\ell+1)$-times differentiable at a. By the definition of $(\ell+1)$-times differentiability at a,

$$\lim_{x \to a} \frac{\|(D^\ell f)(x) - (D^\ell f)(a) - (D^{\ell+1} f)(a)(x-a)\|_{\mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n, \cdots, \mathscr{B}(\mathbb{R}^n, \mathbb{R}) \cdots))}}{\|x-a\|_{\mathbb{R}^n}} = 0.$$

Since

$$\begin{aligned}&\Big|\big[(D^\ell f)(x) - (D^\ell f)(a) - (D^{\ell+1} f)(a)(x-a)\big](e_{j_\ell}) \cdots (e_{j_2})(e_{j_1})\Big| \\ &\quad \leqslant \Big\|\big[(D^\ell f)(x) - (D^\ell f)(a) - (D^{\ell+1} f)(a)(x-a)\big](e_{j_\ell}) \cdots (e_{j_2})\Big\|_{\mathscr{B}(\mathbb{R}^n, \mathbb{R})} \|e_{j_1}\|_{\mathbb{R}^n} \\ &\quad \leqslant \big\|(D^\ell f)(x) - (D^\ell f)(a) - (D^{\ell+1} f)(a)(x-a)\big\|_{\mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n, \cdots, \mathscr{B}(\mathbb{R}^n, \mathbb{R}) \cdots))} \|e_{j_1}\|_{\mathbb{R}^n} \cdots \|e_{j_\ell}\|_{\mathbb{R}^n} \\ &\quad = \big\|(D^\ell f)(x) - (D^\ell f)(a) - (D^{\ell+1} f)(a)(x-a)\big\|_{\mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n, \cdots, \mathscr{B}(\mathbb{R}^n, \mathbb{R}) \cdots))},\end{aligned}$$

using (6.6.2) (for the case $k = \ell$) we conclude that

$$\lim_{x \to a} \frac{\left|\frac{\partial^\ell f}{\partial x_{j_\ell} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(x) - \frac{\partial^\ell f}{\partial x_{j_\ell} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(a) - (D^{\ell+1} f)(a)(\mathrm{e}_{j_1}, \cdots, \mathrm{e}_{j_\ell}, x-a)\right|}{\|x-a\|_{\mathbb{R}^n}}$$
$$= \lim_{x \to a} \frac{\left|(D^\ell f)(x)(\mathrm{e}_{j_1}, \cdots, \mathrm{e}_{j_\ell}) - (D^\ell f)(a)(\mathrm{e}_{j_1}, \cdots, \mathrm{e}_{j_\ell}) - (D^{\ell+1} f)(a)(x-a)(\mathrm{e}_{j_1}, \cdots, \mathrm{e}_{j_\ell})\right|}{\|x-a\|_{\mathbb{R}^n}}$$
$$\leqslant \lim_{x \to a} \frac{\left\|(D^\ell f)(x) - (D^\ell f)(a) - (D^{\ell+1} f)(a)(x-a)\right\|_{\mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n, \cdots, \mathscr{B}(\mathbb{R}^n, \mathbb{R}) \cdots))}}{\|x-a\|_{\mathbb{R}^n}} = 0\,.$$

In particular, if $x = a + t\mathrm{e}_{j_{\ell+1}}$ for some $j_{\ell+1} = 1, \cdots, n$, by the definition of partial derivatives we conclude that

$$(D^{\ell+1} f)(a)(\mathrm{e}_{j_1}, \cdots, \mathrm{e}_{j_\ell}, \mathrm{e}_{j_{\ell+1}}) = \lim_{t \to 0} \frac{\frac{\partial^\ell f}{\partial x_{j_\ell} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(a + t\mathrm{e}_{j_{\ell+1}}) - \frac{\partial^\ell f}{\partial x_{j_\ell} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(a)}{t}$$
$$= \frac{\partial^{\ell+1} f}{\partial x_{j_{\ell+1}} \partial x_{j_\ell} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}(a)$$

which is (6.6.2) for the case $k = \ell + 1$. ▫

Example 6.75. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x_1, x_2) = x_1^2 \cos x_2$, and $u^{(1)} = (2, 0)$, $u^{(2)} = (1, 1)$, $u^{(3)} = (0, -1)$. Suppose that f is three-times differentiable at $a = (0, 0)$ (in fact it is, but we have not talked about this yet). Then

$$(D^3 f)(a)(u^{(1)}, u^{(2)}, u^{(3)}) = \sum_{i,j,k=1}^{2} \frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}(a) u_i^{(1)} u_j^{(2)} u_k^{(3)} = \sum_{j=1}^{2} \frac{\partial^3 f}{\partial x_2 \partial x_j \partial x_1}(a) \cdot 2 \cdot u_j^{(2)} \cdot (-1)$$
$$= \frac{\partial^3 f}{\partial x_2 \partial x_1^2}(0,0) \cdot 2 \cdot 1 \cdot (-1) + \frac{\partial^3 f}{\partial x_2^2 \partial x_1}(0,0) \cdot 2 \cdot 1 \cdot (-1) = 0\,.$$

Corollary 6.76. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$ be $(k+1)$-times differentiable at a. Then for $u^{(1)}, \cdots, u^{(k)}, u^{(k+1)} \in \mathbb{R}^n$,*

$$(D^{k+1} f)(a)(u^{(1)}, \cdots, u^{(k)}, u^{(k+1)}) = \sum_{j=1}^{n} u_j^{(k+1)} \frac{\partial}{\partial x_j}\Big|_{x=a} (D^k f)(x)(u^{(1)}, \cdots, u^{(k)})\,.$$

In other words, (using the terminology in Remark 6.58) $(D^{k+1} f)(a)(u^{(1)}, \cdots, u^{(k)}, u^{(k+1)})$ is the "directional derivative" of the function $(D^k f)(\cdot)(u^{(1)}, \cdots, u^{(k)})$ at a in the "direction" $u^{(k+1)}$.

Proof. By Proposition 6.74,

$$
\begin{aligned}
(D^{k+1}f)(a)(u^{(1)},\cdots,u^{(k)},u^{(k+1)}) &= \sum_{j_1,\cdots,j_k,j_{k+1}=1}^{n} \frac{\partial^{k+1} f}{\partial x_{j_{k+1}}\partial x_{j_k}\cdots\partial x_{j_1}}(a)u^{(1)}_{j_1}\cdots u^{(k)}_{j_k}u^{(k+1)}_{j_{k+1}} \\
&= \sum_{j_{k+1}=1}^{n} u^{(k+1)}_{j_{k+1}} \sum_{j_1,\cdots,j_k=1}^{n} \frac{\partial^{k+1} f}{\partial x_{j_{k+1}}\partial x_{j_k}\cdots\partial x_{j_1}}(a)u^{(1)}_{j_1}\cdots u^{(k)}_{j_k} \\
&= \sum_{j_{k+1}=1}^{n} u^{(k+1)}_{j_{k+1}} \frac{\partial}{\partial x_{j_{k+1}}}\Big|_{x=a} \sum_{j_1,\cdots,j_k=1}^{n} \frac{\partial^{k} f}{\partial x_{j_k}\cdots\partial x_{j_1}}(x)u^{(1)}_{j_1}\cdots u^{(k)}_{j_k} \\
&= \sum_{j_{k+1}=1}^{n} u^{(k+1)}_{j_{k+1}} \frac{\partial}{\partial x_{j_{k+1}}}\Big|_{x=a} (D^k f)(x)(u^{(1)},\cdots,u^{(k)}) .
\end{aligned}
$$

□

Example 6.77. Let $f:\mathbb{R}^2\to\mathbb{R}$ be twice differentiable at $a=(a_1,a_2)\in\mathbb{R}^2$. Then the proposition above suggests that for $u=(u_1,u_2), v=(v_1,v_2)\in\mathbb{R}^2$,

$$
\begin{aligned}
(D^2f)(a)(v)(u) = (D^2f)(a)(u,v) &= \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_j \partial x_i}(a)u_i v_j \\
&= \frac{\partial^2 f}{\partial x_1^2}(a)u_1v_1 + \frac{\partial^2 f}{\partial x_2\partial x_1}(a)u_1v_2 + \frac{\partial^2 f}{\partial x_1\partial x_2}(a)u_2v_1 + \frac{\partial^2 f}{\partial x_2^2}(a)u_2v_2 \\
&= \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(a) & \dfrac{\partial^2 f}{\partial x_2\partial x_1}(a) \\ \dfrac{\partial^2 f}{\partial x_1\partial x_2}(a) & \dfrac{\partial^2 f}{\partial x_2^2}(a) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} .
\end{aligned}
$$

In general, if $f:\mathbb{R}^n\to\mathbb{R}$ be twice differentiable at $a=(a_1,\cdots,a_n)\in\mathbb{R}^n$. Then for $u=(u_1,\cdots,u_n), v=(v_1,\cdots,v_n)\in\mathbb{R}^2$

$$
(D^2f)(a)(v)(u) = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(a) & \cdots & \dfrac{\partial^2 f}{\partial x_n\partial x_1}(a) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1\partial x_n}(a) & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}(a) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} .
$$

The bilinear form $B:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ given by

$$
B(u,v) = (D^2f)(a)(v)(u) \qquad \forall\, u,v\in\mathbb{R}^n
$$

is called the ***Hessian*** of f, and is represented (in the matrix form) as an $n \times n$ matrix by

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{bmatrix}.$$

If the second partial derivatives $\frac{\partial^2 f}{\partial x_j \partial x_i}(a)$ of f at a exists for all $i, j = 1, \cdots, n$ (here the twice differentiability of f at a is ignored), the matrix (on the right-hand side of equality) above is also called the ***Hessian matrix*** of f at a.

Even though there is no reason to believe that $(D^2 f)(a)(u, v) = (D^2 f)(a)(v, u)$ (since the left-hand side means first differentiating f in u-direction and then differentiating Df in v-direction, while the right-hand side means first differentiating f in v-direction then differentiating Df in u-direction), it is still reasonable to ask whether $(D^2 f)(a)$ is symmetric or not; that is, could it be true that $(D^2 f)(a)(u, v) = (D^2 f)(a)(v, u)$ for all $u, v \in \mathbb{R}^n$? When f is twice differentiable at a, this is equivalent of asking (by plugging in $u = \mathrm{e}_i$ and $v = \mathrm{e}_j$) that whether or not

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) . \tag{6.6.3}$$

The following example provides a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that (6.6.3) does not hold at $a = (0, 0)$. We remark that the function in the following example is not twice differentiable at a even though the Hessian matrix of f at a can still be computed.

Example 6.78. 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

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

and

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

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

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

while

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

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

Definition 6.79. A function is said to be ***of class*** $\mathscr{C}^r$ if the first r derivatives exist and are continuous. A function is said to be ***smooth*** or ***of class*** $\mathscr{C}^\infty$ if it is of class $\mathscr{C}^r$ for all positive integer r.

Now we would like to answer the question of what kind of functions are k-times differentiable. Suppose that $\mathcal{U} \subseteq \mathbb{R}^n$ is open and $f : \mathcal{U} \to \mathbb{R}$. Note that by the definition of differentiability, f is k-times differentiable in $\mathcal{U}$ if and only if $D^{k-1}f$ is differentiable in $\mathcal{U}$. This would further imply that f is k-times differentiable in $\mathcal{U}$ if and only if $D^{k-2}f$ is twice differentiable in $\mathcal{U}$. Therefore, Proposition 6.27 and Corollary 6.32 imply that

$$\begin{aligned}
&f \text{ is } k\text{-times (continuously) differentiable in } \mathcal{U}\\
&\quad\Leftrightarrow Df \text{ is } (k-1)\text{-times (continuously) differentiable in } \mathcal{U}\\
&\quad\Leftrightarrow \Big[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\Big] \text{ is } (k-1)\text{-times (continuously) differentiable in } \mathcal{U}\\
&\quad\Leftrightarrow \frac{\partial f}{\partial x_{j_1}} \text{ is } (k-1)\text{-times (continuously) differentiable in } \mathcal{U} \text{ for all } 1 \leqslant j_1 \leqslant n\\
&\quad\Leftrightarrow D\frac{\partial f}{\partial x_{j_1}} \text{ is } (k-2)\text{-times (continuously) differentiable in } \mathcal{U} \text{ for all } 1 \leqslant j_1 \leqslant n\\
&\quad\Leftrightarrow \Big[\frac{\partial^2 f}{\partial x_1 \partial x_{j_1}}, \cdots, \frac{\partial^2 f}{\partial x_n \partial x_{j_1}}\Big] \text{ is } (k-2)\text{-times (continuously) differentiable in } \mathcal{U}\\
&\qquad\ \text{for all } 1 \leqslant j_1 \leqslant n\\
&\quad\Leftrightarrow \frac{\partial^2 f}{\partial x_{j_2} \partial x_{j_1}} \text{ is } (k-2)\text{-times (continuously) differentiable in } \mathcal{U} \text{ for all } 1 \leqslant j_1, j_2 \leqslant n\,.
\end{aligned}$$

Applying similar argument several times, we obtain the following theorem which is an analogy of Corollary 6.32.

Theorem 6.80. *Let $\mathcal{U} \to \mathbb{R}^n$ and $f : \mathcal{U} \to \mathbb{R}$. Suppose that the partial derivative $\dfrac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}$ exists in a neighborhood of $a \in \mathcal{U}$ and is continuous at a for all $j_1, \cdots, j_k =$*

$1, \cdots, n$. *Then* f *is* k*-times differentiable at* a. *Moreover, if* $\dfrac{\partial^k f}{\partial x_{j_k} \partial x_{j_{k-1}} \cdots \partial x_{j_1}}$ *is continuous on* $\mathcal{U}$, *then* f *is of class* $\mathscr{C}^k$.

Theorem 6.81. *Let* $\mathcal{U} \subseteq \mathbb{R}^n$ *be open, and* $f : \mathcal{U} \to \mathbb{R}$. *Suppose that the mixed partial derivatives* $\dfrac{\partial f}{\partial x_i}, \dfrac{\partial f}{\partial x_j}, \dfrac{\partial^2 f}{\partial x_j \partial x_i}, \dfrac{\partial^2 f}{\partial x_j \partial x_i}$ *exist in a neighborhood of* a, *and are continuous at* a. *Then*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a)\,. \tag{6.6.4}$$

Proof. Let $S(a, h, k) = f(a + h\mathrm{e}_i + k\mathrm{e}_j) - f(a + h\mathrm{e}_i) - f(a + k\mathrm{e}_j) + f(a)$, and define $\varphi(x) = f(x + h\mathrm{e}_i) - f(x)$ as well as $\psi(x) = f(x + k\mathrm{e}_j) - f(x)$ for x in a neighborhood of a. Then $S(a, h, k) = \varphi(a+k\mathrm{e}_j) - \varphi(a) = \psi(a+h\mathrm{e}_i) - \psi(a)$; thus the mean value theorem implies that there exists c on the line segment joining a and $a + k\mathrm{e}_j$ and d on the line segment joining a and $a + h\mathrm{e}_i$ such that

$$S(a, h, k) = \varphi(a + k\mathrm{e}_j) - \varphi(a) = k\frac{\partial \varphi}{\partial x_j}(c) = k\Big(\frac{\partial f}{\partial x_j}(c + h\mathrm{e}_i) - \frac{\partial f}{\partial x_j}(c)\Big)\,,$$
$$S(a, h, k) = \psi(a + h\mathrm{e}_i) - \psi(a) = h\frac{\partial \psi}{\partial x_i}(d) = h\Big(\frac{\partial f}{\partial x_i}(d + k\mathrm{e}_j) - \frac{\partial f}{\partial x_i}(d)\Big)\,.$$

As a consequence, if $h \neq 0 \neq k$,

$$\frac{1}{k}\Big(\frac{\partial f}{\partial x_i}(d + k\mathrm{e}_j) - \frac{\partial f}{\partial x_i}(d)\Big) = \frac{S(a, h, k)}{hk} = \frac{1}{h}\Big(\frac{\partial f}{\partial x_j}(c + h\mathrm{e}_i) - \frac{\partial f}{\partial x_j}(c)\Big)$$

By the mean value theorem again, there exists c_1 and d_1 on the line segment joining c, $c + h\mathrm{e}_i$ and d, $d + k\mathrm{e}_j$, respectively, such that

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(d_1) = \frac{\partial^2 f}{\partial x_i \partial x_j}(c_1)\,.$$

The theorem is then concluded by the continuity of $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$ and $\dfrac{\partial^2 f}{\partial x_j \partial x_i}$ at a, and $c_1 \to a$ and $d_1 \to a$ as $(h, k) \to (0, 0)$. ◻

Corollary 6.82. *Let* $\mathcal{U} \subseteq \mathbb{R}^n$ *be open, and* f *is of class* $\mathscr{C}^2$. *Then*

$$(D^2 f)(a)(u, v) = (D^2 f)(a)(v, u) \qquad \forall\, a \in \mathcal{U} \text{ and } u, v \in \mathbb{R}^n\,.$$

Remark 6.83. In view of Remark 6.70, (6.6.4) is the same as the following identity

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

which implies that the order of the two limits $\lim\limits_{h\to 0}$ and $\lim\limits_{k\to 0}$ can be interchanged without changing the value of the limit (under certain conditions).

Example 6.84. Let $f(x,y) = yx^2\cos y^2$. Then

$$\begin{aligned} f_{xy}(x,y) &= (2xy\cos y^2)_y = 2x\cos y^2 - 2xy(2y)\sin y^2 = 2x\cos y^2 - 4xy^2\sin y^2\,, \\ f_{yx}(x,y) &= (x^2\cos y^2 - yx^2(2y)\sin y^2)_x = (x^2\cos y^2 - 2x^2y^2\sin y^2)_x \\ &= 2x\cos y^2 - 4xy^2\sin y^2 = f_{xy}(x,y)\,. \end{aligned}$$

The following two theorems concern the $\mathscr{C}^k$-regularity of inverse functions and implicit functions.

Theorem 6.85. *Let $\mathcal{D} \subseteq \mathbb{R}^n$ be open, $f : \mathcal{D} \to \mathbb{R}^n$ be injective and be of class $\mathscr{C}^k$. If f^{-1}, the inverse function of f, exists and is differentiable in $f(\mathcal{D})$, then f^{-1} is of class $\mathscr{C}^k$.*

Proof. Let $y_0 \in f(\mathcal{D})$. Then $y_0 = f(x_0)$ for some $x_0 \in \mathcal{D}$. Since f is differentiable at x_0 and f^{-1} is differentiable at y_0, by the chain rule we must have

$$\mathrm{I}_n = [D(f\circ f^{-1})](y_0) = [Df](x_0)[Df^{-1}](y_0)\,,$$

where I_n is the $n\times n$ identity matrix. Therefore, $[Df](x_0)$ is invertible, and the inverse function theorem implies that f^{-1} is of class $\mathscr{C}^1$ (in a neighborhood of y_0).

We note that the map $g : \mathrm{GL}(n) \to \mathrm{GL}(n)$ given by $g(L) = L^{-1}$ is infinitely many times differentiable; thus using the identity (from the inverse function theorem)

$$(Df^{-1})(y) = \big((Df)(x)\big)^{-1} = \big(g\circ(Df)\circ f^{-1}\big)(y)\,,$$

by the chain rule we find that if $f \in \mathscr{C}^k$, then $Df^{-1} \in \mathscr{C}^{k-1}$ which is the same as saying that $f^{-1} \in \mathscr{C}^k$. □

Theorem 6.86. *Let $\mathcal{D} \subseteq \mathbb{R}^n\times\mathbb{R}^m$ be open, and $F : \mathcal{D} \to \mathbb{R}^m$ be a function of class $\mathscr{C}^k$. Suppose that for some open set $\mathcal{U} \subseteq \mathbb{R}^m$ and some differentiable function $f : \mathcal{U} \to \mathbb{R}^m$, $\mathcal{U}\times f(\mathcal{U}) \subseteq \mathcal{D}$ and $F(x, f(x)) = 0$ for all $x \in \mathcal{U}$. Then f is of class $\mathscr{C}^k$.*

Proof. **(Not yet finished!!!)** □

6.7 Taylor's Theorem

Recall the Taylor theorem for functions of one variable that if $f : (a,b) \to \mathbb{R}$ be of class $\mathscr{C}^{k+1}$ for some $k \in \mathbb{N}$ and $c \in (a,b)$, then for all $x \in (a,b)$, there exists d in between c and x such that

$$f(x) = \sum_{j=0}^{k} \frac{f^{(j)}(c)}{j!}(x-c)^j + \frac{f^{(k+1)}(d)}{(k+1)!}(x-c)^{k+1},$$

where $f^{(j)}(c)$ denotes the j-th derivative of f at c. In this section, we extend this result to functions of several variables.

Theorem 6.87 (Taylor). *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$ be $(k+1)$-times differentiable. Suppose that $x, a \in \mathcal{U}$ and the line segment joining x and a lies in $\mathcal{U}$. Then there exists a point c on the line segment joining x and a such that*

$$\begin{aligned} f(x) - f(a) = &\sum_{j=1}^{k} \frac{1}{j!}(D^j f)(a)(\overbrace{x-a, \cdots, x-a}^{j \text{ copies of } x-a}) \\ &+ \frac{1}{(k+1)!}(D^{k+1} f)(c)(\underbrace{x-a, \cdots, x-a}_{(k+1) \text{ copies of } x-a}) . \end{aligned} \tag{6.7.1}$$

Proof. Let $g(t) = f\big((1-t)a + tx\big)$. Since $\overline{xa} \subseteq \mathcal{U}$ and $\mathcal{U}$ is open, there exists $\delta > 0$ such that $(1-t)a + tx \in \mathcal{U}$ for all $t \in (-\delta, 1+\delta)$. By the chain rule, for $t \in (-\delta, 1+\delta)$,

$$g'(t) = (Df)\big((1-t)a + tx\big)(x-a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\big((1-t)a + tx\big)(x_i - a_i)\,;$$

thus for $t \in (-\delta, 1+\delta)$, Proposition 6.74 shows that

$$g''(t) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}\big((1-t)a + tx\big)(x_i - a_i)(x_j - a_j) = (D^2 f)\big((1-t)a + tx\big)(x-a, x-a).$$

By induction, we conclude that

$$g^{(j)}(t) = (D^j f)\big((1-t)a + tx\big)(\underbrace{x-a, \cdots, x-a}_{j \text{ copies of } x-a}) .$$

By the fact that f is $(k+1)$-times differentiable, $g : (-\delta, 1+\delta) \to \mathbb{R}$ is $(k+1)$-times differentiable as well. Theorem 4.68 then implies that for some $t_0 \in (0,1)$,

$$g(1) - g(0) = \sum_{j=1}^{k} \frac{g^{(j)}(0)}{j!} + \frac{g^{(k+1)}(t_0)}{(k+1)!} . \tag{6.7.2}$$

Letting $c = (1-t_0)a + t_0 x$, (6.7.2) implies (6.7.1). □

Definition 6.88. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$ be k-times differentiable. The k-th degree Taylor polynomial for f centered at a is the polynomial

$$\sum_{j=0}^{k} \frac{1}{j!}(D^j f)(a)(\underbrace{x-a, \cdots, x-a}_{j \text{ copies } x-a}).$$

Corollary 6.89. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $f : \mathcal{U} \to \mathbb{R}$ be $(k+1)$-times differentiable, and define the remainder*

$$R_k(a,h) = f(a+h) - \sum_{j=0}^{k} \frac{1}{j!}(D^j f)(a)(h, \cdots, h).$$

Then $\displaystyle\lim_{h\to 0} \frac{R_k(a,h)}{\|h\|^k_{\mathbb{R}^n}} = 0$, or in notation, $R_k(a,h) = o(\|h\|^k_{\mathbb{R}^n})$ as $h \to 0$.

Example 6.90. Let $f(x,y) = e^x \cos y$. Compute the fourth degree Taylor polynomial for f centered at $(0,0)$.

Solution: We compute the zeroth, the first, the second, the third and the fourth mixed derivatives of f at $(0,0)$ as follows:

$$\begin{aligned}
f(0,0) &= 1\,, \qquad f_x(0,0) = 1\,, \qquad f_y(0,0) = 0\,,\\
f_{xx}(0,0) &= 1\,, \qquad f_{xy}(0,0) = f_{yx}(0,0) = 0\,, \qquad f_{yy}(0,0) = -1\,,\\
f_{xxx}(0,0) &= 1\,, \qquad f_{xxy}(0,0) = f_{xyx}(0,0) = f_{yxx}(0,0) = 0\,,\\
f_{yyy}(0,0) &= 0\,, \qquad f_{yyx}(0,0) = f_{yxy}(0,0) = f_{xyy}(0,0) = -1\,,
\end{aligned}$$

and

$$\begin{aligned}
&f_{xxxx}(0,0) = 1\,, \qquad f_{yyyy}(0,0) = 1\,,\\
&f_{xxxy}(0,0) = f_{xxyx}(0,0) = f_{xyxx}(0,0) = f_{yxxx}(0,0) = 0\,,\\
&f_{xyyy}(0,0) = f_{yxyy}(0,0) = f_{yyxy}(0,0) = f_{yyyx}(0,0) = 0\,,\\
&f_{xxyy}(0,0) = f_{xyxy}(0,0) = f_{xyyx}(0,0) = f_{yxxy}(0,0)\\
&\qquad\qquad = f_{yxyx}(0,0) = f_{yyxx}(0,0) = -1\,.
\end{aligned}$$

Then the fourth degree Taylor polynomial for f centered at $(0,0)$ is

$$\begin{aligned} f(0,0) &+ f_x(0,0)x + f_y(0,0)y + \frac{1}{2}\Big[f_{xx}(0,0)x^2 + 2f_{xy}(0,0)xy + f_{yy}(0,0)y^2\Big] \\ &+ \frac{1}{6}\Big[f_{xxx}(0,0)x^3 + 3f_{xxy}(0,0)x^2y + 3f_{xyy}(0,0)xy^2 + f_{yyy}(0,0)y^3\Big] \\ &+ \frac{1}{24}\Big[f_{xxxx}(0,0)x^4 + 4f_{xxxy}(0,0)x^3 + 6f_{xxyy}(0,0)x^2y^2 \\ &\qquad + 4f_{xyyy}(0,0)xy^3 + f_{yyyy}(0,0)y^4\Big] \\ &= 1 + x + \frac{1}{2}\big(x^2 - y^2\big) + \frac{1}{6}\big(x^3 - 3xy^2\big) + \frac{1}{24}\big(x^4 - 6x^2y^2 + y^4\big)\,. \end{aligned}$$

Observing that using the Taylor expansions

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \cdots \quad \text{and} \quad \cos y = 1 - \frac{1}{2}y^2 + \frac{1}{24}y^4 + \cdots\,,$$

we can "formally" compute $e^x \cos y$ by multiplying the two "polynomials" above and obtain that

$$e^x \cos y \text{ ``='' } 1 + x + \frac{1}{2}\big(x^2 - y^2\big) + \Big(\frac{1}{6}x^3 - \frac{1}{2}xy^2\Big) + \Big(\frac{1}{24}x^4 - \frac{1}{4}x^2y^2 + \frac{1}{24}y^2\Big) + \text{h.o.t.}\,;$$

where h.o.t. stands for the higher order terms which are terms with fifth or higher degree.

Definition 6.91. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open. A function $f : \mathcal{U} \to \mathbb{R}$ is said to be ***real analytic*** at $a \in \mathcal{U}$ if $f(x) = \sum\limits_{k=0}^{\infty} \frac{1}{k!}(D^k f)(a)(x-a, \cdots, x-a)$ in a neighborhood of a.

Example 6.92. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \exp\big(-\dfrac{1}{|x|^2}\big) & \text{if } x > 0\,, \\ 0 & \text{if } x < 0\,. \end{cases}$$

Then f is of class $\mathscr{C}^\infty$, and $f^{(k)}(0) = 0$ for all $k \in \mathbb{N}$. Therefore, f is not real analytic at 0.

6.8 Maxima and Minima

Definition 6.93. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$.

1. If there is a neighborhood of $x_0 \in \mathcal{U}$ such that $f(x_0)$ is a maximum in this neighborhood, then x_0 is called a ***local maximum point*** of f.

2. If there is a neighborhood of $x_0 \in \mathcal{U}$ such that $f(x_0)$ is a minimum in this neighborhood, then x_0 is called a ***local minimum point*** of f.

3. A point is called an ***extreme point*** of f if it is either a local maximum point or a local minimum point of f.

4. A point x_0 is a ***critical point*** of f if f is differentiable at x_0 and $(Df)(x_0) = 0$; that is, $(Df)(x_0) \in \mathscr{B}(\mathbb{R}^n, \mathbb{R})$ is the trivial map (which sends every vector in $\mathbb{R}^n$ to zero vector).

5. A point x_0 is a ***saddle point*** of f if x_0 is a critical point of f but not an extreme point of f.

Theorem 6.94. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, $f : \mathcal{U} \to \mathbb{R}$ be differentiable, and $x_0 \in \mathcal{U}$ is an extreme point of f. Then x_0 is a critical point of f.*

Proof. Suppose the contrary that the linear map $(Df)(x_0) : \mathbb{R}^n \to \mathbb{R}$ is not the zero map; that is, there exists $u \in \mathbb{R}^n$, $u \neq 0$, such that $(Df)(x_0)(u) = c \neq 0$ for some constant $c \in \mathbb{R}$. W.L.O.G, we can assume that $\|u\|_{\mathbb{R}^n} = 1$ and $c > 0$ (for otherwise change u to $-u$). By the differentiability of f,

$$\exists\, \delta > 0 \ni \big|f(x_0 + h) - f(x_0) - (Df)(x_0)(h)\big| \leqslant \frac{c}{2}\|h\|_{\mathbb{R}^n} \quad \text{whenever} \quad \|h\|_{\mathbb{R}^n} < \delta\,.$$

Then for any $0 < \lambda < \delta$,

$$\big|f(x_0 \pm \lambda u) - f(x_0) \mp \lambda (Df)(x_0)(u)\big| \leqslant \frac{\lambda c}{2}\,.$$

Therefore, $-\frac{\lambda c}{2} \leqslant f(x_0 \pm \lambda u) - f(x_0) \mp \lambda c \leqslant \frac{\lambda c}{2}$ which further implies that

$$f(x_0) \leqslant f(x_0 + \lambda u) - \frac{\lambda c}{2} < f(x_0 + \lambda u) \ \text{ and } \ f(x_0) \geqslant f(x_0 - \lambda u) + \frac{\lambda c}{2} > f(x_0 - \lambda u)$$

for all $\lambda > 0$ small enough. As a consequence, x_0 cannot be a local extreme point of f, a contradiction. □

Definition 6.95. If $f : \mathcal{U} \to \mathbb{R}$ is of class $\mathscr{C}^2$, the ***Hessian of f at*** x_0 is the bilinear function $H_{x_0}(f) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$H_{x_0}(f)(u, v) = (D^2 f)(x_0)(u, v) \qquad \forall\, u, v \in \mathbb{R}^n\,.$$

The matrix representation of $H_{x_0}(f)(\cdot,\cdot)$ is given by

$$\big[H_{x_0}(f)\big] = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(x_0) & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_1}(x_0) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x_0) & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}(x_0) \end{bmatrix}$$

in the sense that $H_{x_0}(f)(u,v) = [u]^{\mathrm{T}}\big[H_{x_0}(f)\big][v] = [v]^{\mathrm{T}}\big[H_{x_0}(f)\big][u]$.

Definition 6.96. A bilinear form $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called ***positive definite*** / ***negative definite*** if $B(u,u) \begin{smallmatrix} > \\ < \end{smallmatrix} 0$ for all $u \neq 0$, and is called ***positive semi-definite*** / ***negative semi-definite*** if $B(u,u) \begin{smallmatrix} \geqslant \\ \leqslant \end{smallmatrix} 0$ for all $u \in \mathbb{R}^n$.

Theorem 6.97. *Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$ be a function of class $\mathscr{C}^2$.*

1. *If x_0 is a critical point of f such that the Hessian $H_{x_0}(f)$ is negative / positive definite, then f has a local maximum / minimum point at x_0.*

2. *If f has a local maximum / minimum point at x_0, then $H_{x_0}(f)$ is negative / positive semi-definite.*

Proof. 1. Suppose that $H_{x_0}(f)$ is negative definite.

Claim: There exists $0 < \lambda < \infty$ such that

$$H_{x_0}(f)(u,u) \leqslant -\lambda \|u\|_{\mathbb{R}^n}^2 \qquad \forall\, u \in \mathbb{R}^n. \tag{6.8.1}$$

Proof of claim: Since $H_{x_0}(f)(u,u)$, viewed as a function of u, is continuous, by Theorem 4.21 $\lambda = -\max\limits_{\|u\|_{\mathbb{R}^n}=1} H_{x_0}(f)(u,u)$ exists and is positive. Then for all $u \in \mathbb{R}^n$ with $u \neq 0$,

$$H_{x_0}(f)\Big(\frac{u}{\|u\|_{\mathbb{R}^n}}, \frac{u}{\|u\|_{\mathbb{R}^n}}\Big) \leqslant -\lambda \qquad \forall\, u \in \mathbb{R}^n, u \neq 0\,.$$

The inequality (6.8.1) follows from that the Hessian $H_{x_0}(f)$ is bilinear.

Since $f \in \mathscr{C}^2$, there exists $\delta > 0$ such that $D(x_0,\delta) \subseteq \mathcal{U}$ and

$$\big\|(D^2 f)(x) - (D^2 f)(x_0)\big\|_{\mathscr{B}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n,\mathbb{R}))} \leqslant \frac{\lambda}{2} \qquad \forall\, x \in D(x_0,\delta). \tag{6.8.2}$$

Now since x_0 is a critical point of f, $(Df)(x_0) = 0$. As a consequence, by Taylor's theorem (Theorem 6.87), for any $x \in D(x_0, \delta)$, we can find $c = c(x) \in \overline{xx_0}$ such that

$$\begin{aligned} f(x) &= f(x_0) + (Df)(x_0)(x - x_0) + \frac{1}{2}(D^2 f)(c)(x - x_0, x - x_0) \\ &= f(x_0) + \frac{1}{2}(D^2 f)(x_0)(x - x_0, x - x_0) + \frac{1}{2}\big[(D^2 f)(c) - (D^2 f)(x_0)\big](x - x_0, x - x_0) \\ &\leqslant f(x_0) - \frac{1}{2}\lambda \|x - x_0\|^2_{\mathbb{R}^n} + \frac{1}{2}\Big|\big[(D^2 f)(c) - (D^2 f)(x_0)\big](x - x_0, x - x_0)\Big| \\ &\leqslant f(x_0) - \frac{1}{2}\lambda \|x - x_0\|^2_{\mathbb{R}^n} + \frac{1}{2}\big\|(D^2 f)(c) - (D^2 f)(x_0)\big\|_{\mathscr{B}(\mathbb{R}^n;\mathscr{B}(\mathbb{R}^n,\mathbb{R}))} \|x - x_0\|^2_{\mathbb{R}^n} . \end{aligned}$$

Note that $c = c(x) \in D(x_0, \delta)$ if $x \in D(x_0, \delta)$; thus (6.8.2) implies that if $x \in D(x_0, \delta)$,

$$f(x) \leqslant f(x_0) - \frac{\lambda}{2}\|x - x_0\|^2_{\mathbb{R}^n} + \frac{1}{2}\frac{\lambda}{2}\|x - x_0\|^2_{\mathbb{R}^n} \leqslant f(x_0) - \frac{\lambda}{4}\|x - x_0\|^2_{\mathbb{R}^n} .$$

As a consequence, for all $x \in D(x_0, \delta)$, $f(x) \leqslant f(x_0)$ which validates that x_0 is a local maximum point of f.

2. Suppose the contrary that f has a local maximum point at x_0 but for some $u \in \mathbb{R}^n$,

$$H_{x_0}(f)(u, u) > 0 .$$

W.L.O.G, we can assume that $\|u\|_{\mathbb{R}^n} = 1$. By Theorem 6.94, $(Df)(x_0) = 0$; thus Taylor's Theorem implies that

$$f(x) = f(x_0) + \frac{1}{2}(D^2 f)(c)(x - x_0, x - x_0) = f(x_0) + \frac{1}{2}(x - x_0)^{\mathrm{T}}\big[H_c(f)\big](x - x_0) .$$

Since x_0 is a local maximum point of f, there exists $\delta > 0$ such that $f(x) \leqslant f(x_0)$ for all $x \in D(x_0, \delta)$. As a consequence, for some $c = c(x) \in \overline{xx_0}$,

$$(x - x_0)^{\mathrm{T}}\big[H_c(f)\big](x - x_0) = 2\big[f(x) - f(x_0)\big] \leqslant 0 \qquad \forall\, x \in D(x_0, \delta) .$$

Let $0 < t < \delta$ and $x = x_0 + tu$. Then $x \in D(x_0, \delta)$; thus

$$H_c(f)(u, u) \leqslant 0 \qquad \forall\, t \in (0, \delta) .$$

We note that c depends on t, and $c \to x_0$ as $t \to 0$. Therefore, by the continuity of $H_\bullet(f)$, passing $t \to 0$ in the inequality above we find that

$$H_{x_0}(f)(u, u) = \lim_{t \to 0} H_c(f)(u, u) \leqslant 0$$

which is a contradiction. □

Remark 6.98. Inequality (6.8.1) can also be obtained by studying the largest eigenvalue of $H_{x_0}(f)$. Note that since $f \in \mathscr{C}^2$, $H_{x_0}(f)$ is symmetric by Theorem 6.81. As a consequence, there exists an orthonormal matrix $\mathbb{O} \in \mathrm{GL}(n)$ whose columns are (real) eigenvectors of $H_{x_0}(f)$

$$\left[H_{x_0}(f)\right] = \mathbb{O}\Lambda\mathbb{O}^{\mathrm{T}} ,$$

where Λ is a diagonal matrix whose diagonal entries are eigenvalues of $H_{x_0}(f)$. Note that by the orthonormality of $\mathbb{O}$, every vector $u \in \mathbb{R}^n$ satisfies $\|\mathbb{O}^{\mathrm{T}}u\|_{\mathbb{R}^n} = \|u\|_{\mathbb{R}^n}$. Therefore,

$$H_{x_0}(f)(u,u) = u^{\mathrm{T}}\mathbb{O}\Lambda\mathbb{O}^{\mathrm{T}}u = (\mathbb{O}^{\mathrm{T}}u)^{\mathrm{T}}\Lambda(\mathbb{O}^{\mathrm{T}}u) \leqslant \lambda\|\mathbb{O}^{\mathrm{T}}u\|^2_{\mathbb{R}^n} = \lambda\|u\|^2_{\mathbb{R}^n} ,$$

where λ is the largest eigenvalue of Λ.

Remark 6.99 (Sylvester's criterion)**.** To justify if a matrix $\left[H_{x_0}(f)\right]$ is positive/negative definite, let

$$\Delta_k = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f}{\partial x_k \partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_k} & \cdots & \dfrac{\partial^2 f}{\partial x_k^2} \end{bmatrix} (x_0) \, .$$

Then $H_{x_0}(f)$ is $\begin{matrix}\text{positive}\\\text{negative}\end{matrix}$ definite if and only if $\begin{matrix}\det(\Delta_k) > 0\\(-1)^k\det(\Delta_k) > 0\end{matrix}$ for all $k = 1, \cdots, n$.

6.9 Exercises

Problem 6.1. Let $\{T_k\}_{k=1}^{\infty} \subseteq \mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)$ be a sequence of bounded linear maps from $\mathbb{R}^n \to \mathbb{R}^m$. Prove that the following three statements are equivalent:

1. $\{T_k\}_{k=1}^{\infty}$ converges pointwise (to a function T);

2. $\lim\limits_{k,\ell\to\infty} \|T_k - T_\ell\|_{\mathscr{B}(\mathbb{R}^n,\mathbb{R}^m)} = 0$;

3. $\{T_k\}_{k=1}^{\infty}$ converges uniformly (to T) on every compact subsets of $\mathbb{R}^n$.

Problem 6.2. Let $\mathscr{P}((0,1)) \subseteq \mathscr{C}_b((0,1);\mathbb{R})$ be the collection of all polynomials defined on $(0,1)$.

1. Show that the operator $\dfrac{d}{dx} : \mathscr{P}((0,1)) \to \mathscr{C}_b((0,1))$ is linear.

2. Show that $\frac{d}{dx} : \big(\mathscr{P}((0,1)), \|\cdot\|_\infty\big) \to \big(\mathscr{C}_b((0,1)), \|\cdot\|_\infty\big)$ is unbounded; that is, show that

$$\sup_{\|p\|_\infty=1} \|p'\|_\infty = \infty\,.$$

§6.2 Definition of Derivatives and the Jacobian Matrices

Problem 6.3. Show that if $f : \mathbb{C} \to \mathbb{R}$ is differentiable at z_0, then $(Df)(z_0) = 0$.
Hint: Show that $\mathscr{B}(\mathbb{C}, \mathbb{R}) = \{0\}$.

Problem 6.4. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

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

Compute $\frac{\partial f}{\partial x}(x,y)$ and $\frac{\partial f}{\partial y}(x,y)$.

Problem 6.5. Investigate the differentiability of

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

Problem 6.6. Investigate the differentiability of

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

Problem 6.7. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

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

Discuss the differentiability of f. Find $(\nabla f)(x,y)$ at points of differentiability.

Problem 6.8. Let $r > 0$ and $\alpha > 1$. Suppose that $f : D(0,r) \to \mathbb{R}$ satisfies $|f(x)| \leqslant \|x\|^\alpha$ for all $x \in D(0,r)$. Show that f is differentiable at 0. What happens if $\alpha = 1$?

Problem 6.9. Suppose that $f, g : \mathbb{R} \to \mathbb{R}^m$ are differentiable at a and there is a $\delta > 0$ such that $g(x) \neq 0$ for all $0 < |x-a| < \delta$. If $f(a) = g(a) = 0$ and $(Dg)(a) \neq 0$, show that

$$\lim_{x\to a} \frac{\|f(x)\|}{\|g(x)\|} = \frac{\|(Df)(a)\|}{\|(Dg)(a)\|}\,.$$

Problem 6.10. Consider the map δ defined in Problem 5.11 in Chapter 5; that is, $\delta : \mathscr{C}([0,1];\mathbb{R}) \to \mathbb{R}$ be defined by $\delta(f) = f(0)$. Show that δ is differentiable. Find $(D\delta)(f)$ (for $f \in \mathscr{C}([0,1];\mathbb{R})$).

Problem 6.11. Let $f : \mathrm{GL}(n) \to \mathrm{GL}(n)$ be given by $f(L) = L^{-1}$. In class we have shown that f is continuous on $\mathrm{GL}(n)$. Show that f is differentiable at each "point" (or more precisely, linear map) of $\mathrm{GL}(n)$.
Hint: In order to show the differentiability of f at $L \in \mathrm{GL}(n)$, we need to figure out what $(Df)(L)$ is. So we need to compute $f(L+h) - f(L)$, where $h \in \mathscr{B}(\mathbb{R}^n, \mathbb{R}^n)$ is a "small" linear map. Compute $(L+h)^{-1} - L^{-1}$ and make a conjecture what $(Df)(L)$ should be.

Problem 6.12. Let $I : \mathscr{C}([0,1];\mathbb{R}) \to \mathbb{R}$ be defined by

$$I(f) = \int_0^1 f(x)^2 \, dx \,.$$

Show that I is differentiable at every "point" $f \in \mathscr{C}([0,1];\mathbb{R})$.
Hint: Figure out what $(DI)(f)$ is by computing $I(f+h) - I(f)$, where $h \in \mathscr{C}([0,1];\mathbb{R})$ is a "small" continuous function.
Remark. A map from a space of functions such as $\mathscr{C}([0,1];\mathbb{R})$ to a scalar field such as $\mathbb{R}$ or $\mathbb{C}$ is usually called a ***functional***. The derivative of a functional I is usually denoted by δI instead of DI.

§6.3 Conditions for Differentiability

§6.4 Properties of Differentiable Functions

Problem 6.13. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$. Suppose that the partial derivatives $\dfrac{\partial f}{\partial x_1}, \cdots, \dfrac{\partial f}{\partial x_n}$ are bounded on $\mathcal{U}$; that is, there exists a real number $M > 0$ such that

$$\Big|\frac{\partial f}{\partial x_j}(x)\Big| \leqslant M \qquad \forall\, x \in \mathcal{U} \text{ and } j = 1, \cdots, n \,.$$

Show that f is continuous on $\mathcal{U}$.
Hint: Mimic the proof of Theorem 6.28.

Problem 6.14. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$. Show that f is differentiable at $a \in \mathcal{U}$ if and only if there exists a vector-valued function $\varepsilon : \mathcal{U} \to \mathbb{R}^n$ such that

$$f(x) - f(a) - \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a)(x_j - a_j) = \varepsilon(x) \cdot (x - a)$$

and $\varepsilon(x) \to 0$ as $x \to a$.

Problem 6.15. Verify the chain rule for

$$u(x,y,z) = xe^y, \quad v(x,y,z) = yz\sin x$$

and

$$f(u,v) = u^2 + v\sin u$$

with $h(x,y,z) = f\big(u(x,y,z), v(x,y,z)\big)$.

Problem 6.16. Let (r,θ,φ) be the spherical coordinate of $\mathbb{R}^3$ so that

$$x = r\cos\theta\sin\varphi, y = r\sin\theta\sin\varphi, z = r\cos\varphi\,.$$

1. Find the Jacobian matrices of the map $(x,y,z) \mapsto (r,\theta,\varphi)$ and the map $(r,\theta,\varphi) \mapsto (x,y,z)$.
2. Suppose that $f(x,y,z)$ is a differential function in $\mathbb{R}^3$. Express $|\nabla f|^2$ in terms of the spherical coordinate.

Problem 6.17. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open and convex, and $f : \mathcal{U} \to \mathbb{R}^m$ be differentiable on $\mathcal{U}$. Show that for each $a, b \in \mathcal{U}$ and vector $v \in \mathbb{R}^m$, there exists c on the line segment joining a and b such that

$$v \cdot \big[f(b) - f(a)\big] = v \cdot D(f)(c)(b-a)\,.$$

Problem 6.18. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and for each $1 \leqslant i, j \leqslant n$, $a_{ij} : \mathcal{U} \to \mathbb{R}$ be differentiable functions. Define $A = [a_{ij}]$ and $J = \det(A)$. Show that

$$\frac{\partial J}{\partial x_k} = \operatorname{tr}\big(\operatorname{Adj}(A)\frac{\partial A}{\partial x_k}\big) \qquad \forall\, 1 \leqslant k \leqslant n\,,$$

where for a square matrix $M = [m_{ij}]$, $\operatorname{tr}(M)$ denotes the trace of M, $\operatorname{Adj}(M)$ denotes the adjoint matrix of M, and $\dfrac{\partial M}{\partial x_k}$ denotes the matrix whose (i,j)-th entry is given by $\dfrac{\partial m_{ij}}{\partial x_k}$.
Hint: Show that

$$\frac{\partial J}{\partial x_k} = \begin{vmatrix} \frac{\partial a_{11}}{\partial x_k} & a_{12} & \cdots & a_{1n} \\ \frac{\partial a_{21}}{\partial x_k} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ \frac{\partial a_{n1}}{\partial x_k} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \frac{\partial a_{12}}{\partial x_k} & a_{13} & \cdots & a_{1n} \\ a_{21} & \frac{\partial a_{22}}{\partial x_k} & a_{23} & \cdots & a_{2n} \\ \vdots & & & & \vdots \\ a_{n1} & \frac{\partial a_{n2}}{\partial x_k} & a_{n3} & \cdots & a_{nn} \end{vmatrix} + \cdots + \begin{vmatrix} a_{11} & \cdots & a_{(n-1)1} & \frac{\partial a_{1n}}{\partial x_k} \\ a_{21} & \cdots & a_{(n-1)2} & \frac{\partial a_{2n}}{\partial x_k} \\ \vdots & & & \vdots \\ a_{n1} & \cdots & a_{(n-1)n} & \frac{\partial a_{n1}}{\partial x_k} \end{vmatrix}$$

and rewrite this identity in the form which is asked to prove. You can also show the differentiation formula by applying the chain rule to the composite function $F \circ g$ of maps $g : \mathcal{U} \to \mathbb{R}^{n^2}$ and $F : \mathbb{R}^{n^2} \to \mathbb{R}$ defined by $g(x) = \big(a_{11}(x), a_{12}(x), \cdots, a_{nn}(x)\big)$ and $F(a_{11}, \cdots, a_{nn}) = \det\big([a_{ij}]\big)$. Check first what $\dfrac{\partial F}{\partial a_{ij}}$ is.

Problem 6.19. Let $\psi : \mathbb{R}^n \to \mathbb{R}^n$ be a differentiable function such that $\dfrac{\partial^2 \psi_k}{\partial x_i \partial x_j} = \dfrac{\partial}{\partial x_i}\big(\dfrac{\partial \psi_k}{\partial x_j}\big)$ exists and is continuous in $\mathbb{R}^n$ for each $1 \leqslant i, j, k \leqslant n$. Suppose that $(D\psi)(x) \in \mathrm{GL}(n)$ for all $x \in \mathbb{R}^n$, and define $A = (D\psi)^{-1}$ (or in terms of their matrix representation, $[A] = [D\psi]^{-1}$). Let $\psi = (\psi_1, \cdots, \psi_n)$ and $[A] = [a_{ij}]$.

1. Show that $\sum\limits_{k=1}^{n} a_{ik} \dfrac{\partial \psi_k}{\partial x_j} = \sum\limits_{k=1}^{n} \dfrac{\partial \psi_i}{\partial x_k} a_{kj} = \delta_{ij}$, where δ_{ij} is the Kronecker delta; that is, $\delta_{ij} = 1$ if $i = j$ or $\delta_{ij} = 0$ if $i \neq j$.

2. Show that for each $1 \leqslant i, j, k \leqslant n$, $a_{ij} : \mathbb{R}^n \to \mathbb{R}$ is differentiable, and
$$\frac{\partial a_{ij}}{\partial x_k} = -\sum_{r,s=1}^{n} a_{ir} \frac{\partial^2 \psi_r}{\partial x_k \partial x_s} a_{sj}\,.$$

Problem 6.20. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open and connected, and $f : \mathcal{U} \to \mathbb{R}$ be a function such that $\dfrac{\partial f}{\partial x_j}(x) = 0$ for all $x \in \mathcal{U}$. Show that f is constant in $\mathcal{U}$.

§6.5 Directional Derivatives and Gradient Vectors

Problem 6.21. Let
$$f(x,y) = \begin{cases} \dfrac{x^3 y}{x^4 + y^2} & \text{if } (x,y) \neq (0,0)\,, \\ 0 & \text{if } (x,y) = (0,0)\,. \end{cases}$$
Show that the directional derivative of f at the origin exists in all directions u, and
$$(D_u f)(0,0) = \Big(\frac{\partial f}{\partial x}(0,0), \frac{\partial f}{\partial y}(0,0)\Big) \cdot u\,.$$

§6.6 Higher Derivatives of Functions

Problem 6.22. Let $f(x,y,z) = (x^2+1)\cos(yz)$, and $a = (0, \frac{\pi}{2}, 1)$, $u = (1,0,0)$, $v = (0,1,0)$ and $w = (2,0,1)$.

1. Compute $(Df)(a)(u)$.

2. Compute $(D^2 f)(a)(v)(u)$.

3. Compute $(D^3 f)(a)(w)(v)(u)$.

Problem 6.23. 1. If $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $g : B \subseteq \mathbb{R}^m \to \mathbb{R}^\ell$ are twice differentiable and $f(A) \subseteq B$, then for $x_0 \in A$, $u, v \in \mathbb{R}^n$, show that

$$\begin{aligned} &D^2(g \circ f)(x_0)(u, v) \\ &\quad = (D^2 g)(f(x_0))\big((Df)(x_0)(u), Df(x_0)(v)\big) + (Dg)(f(x_0))\big((D^2 f)(x_0)(u, v)\big)\,. \end{aligned}$$

2. If $p : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map plus some constant; that is, $p(x) = Lx + c$ for some $L \in \mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)$, and $f : A \subseteq \mathbb{R}^m \to \mathbb{R}^s$ is k-times differentiable, prove that

$$D^k(f \circ p)(x_0)(u^{(1)}, \cdots, u^{(k)}) = (D^k f)\big(p(x_0)\big)\big((Dp)(x_0)(u^{(1)}), \cdots, (Dp)(x_0)(u^{(k)}\big)\,.$$

§6.7 Taylor's Theorem

Problem 6.24. Let $f(x, y)$ be a real-valued function on $\mathbb{R}^2$. Suppose that f is of class $\mathscr{C}^1$ (that is, all first partial derivatives are continuous on $\mathbb{R}^2$) and $\dfrac{\partial^2 f}{\partial x \partial y}$ exists and is continuous. Show that $\dfrac{\partial^2 f}{\partial y \partial x}$ exists and $\dfrac{\partial^2 f}{\partial x \partial y} = \dfrac{\partial^2 f}{\partial y \partial x}$.

Hint: Mimic the proof of Theorem 6.81.

Problem 6.25. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable, and Df is a constant map in $\mathscr{B}(\mathbb{R}^n, \mathbb{R}^m)$; that is, $(Df)(x_1)(u) = (Df)(x_2)(u)$ for all $x_1, x_2 \in \mathbb{R}^n$ and $u \in \mathbb{R}^n$. Show that f is a linear term plus a constant and that the linear part of f is the constant value of Df.

Problem 6.26. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}^n$ be of class $\mathscr{C}^2$ such that $Df : \mathcal{U} \to \mathscr{B}(\mathbb{R}^n, \mathbb{R}^n)$ satisfies $(Df)(x) \in \mathrm{GL}(n)$ for all $x \in \mathcal{U}$. Define $J = \det([Df])$ and $A = [Df]^{-1}$. With a_{ij} denoting the (i, j)-th entry of A, show the Piola identity

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(J a_{ij})(x) = 0 \qquad \forall\, 1 \leqslant j \leqslant n \text{ and } x \in \mathcal{U}\,. \tag{6.9.1}$$

Is f continuous at $(0, 0)$? Is f differentiable at $(0, 0)$?

Problem 6.27. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{U} \to \mathbb{R}$ be of class $\mathscr{C}^k$ and $(D^j f)(x_0) = 0$ for $j = 1, \cdots, k-1$, but $(D^k f)(x_0)(u, u, \cdots, u) < 0$ for all $u \in \mathbb{R}^n$, $u \neq 0$. Show that f has a local maximum at x_0; that is, $\exists\, \delta > 0$ such that

$$f(x) \leqslant f(x_0) \qquad \forall\, x \in D(x_0, \delta)\,.$$

§6.8 Maxima and Minima

Problem 6.28. Let $f(x,y) = x^3 + x - 4xy + 2y^2$,

1. Find all critical points of f.
2. Find the corresponding quadratic from $Q(x,y,h,k)$ $\big($or $(D^2 f(x,y)\big((h,k),(h,k)\big)\big)$ at these critical points, and determine which of them is positive definite.
3. Find all relative extrema and saddle points.
4. Find the maximal value of f on the set

$$A = \big\{(x,y) \,\big|\, 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1, x + y \leqslant 1\big\}.$$

Problem 6.29. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x,y) = \begin{cases} x^2 + y^2 - 2x^2 y - \dfrac{4x^6 y^2}{(x^4 + y^2)^2} & \text{if } (x,y) \neq (0,0)\,, \\ 0 & \text{if } (x,y) = (0,0)\,. \end{cases}$$

1. Show that f is continuous (at $(0,0)$) by showing that for all $(x,y) \in \mathbb{R}^2$,

$$4x^4 y^2 \leqslant (x^4 + y^2)^2\,.$$

2. For $0 \leqslant \theta \leqslant 2\pi$, $-\infty < t < \infty$, define

$$g_\theta(t) = f(t\cos\theta, t\sin\theta)\,.$$

Show that each g_θ has a strict local minimum at $t = 0$. In other words, the restriction of f to each straight line through $(0,0)$ has a strict local minimum at $(0,0)$.

3. Show that $(0,0)$ is not a local minimum for f.

Problem 6.30 (True or False). Determine whether the following statements are true or false. If it is true, prove it. Otherwise, give a counter-example.

1.

2.

3.

4. Let $\mathcal{U} \subseteq \mathbb{R}^n$ be open. Then $f : \mathcal{U} \to \mathbb{R}$ is differentiable at $a \in \mathcal{U}$ if and only if each directional derivative $(D_u f)(a)$ exists and

$$(D_u f)(a) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) u_j = \Big(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a)\Big) \cdot u$$

where $u = (u_1, \cdots, u_n)$ is a unit vector.

5. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be of class $\mathscr{C}^1$. Assume that all second order partial derivatives of f exist, then f is second times differentiable in $\mathbb{R}^2$.

6. Let f be a function defined on $\mathbb{R}^2$, and A be an invertible matrix. Define $y = Ax$ for $x \in \mathbb{R}^n$. Then $f(y)$ is differentiable if and only if $f(Ax)$ is differentiable as a function of x.

7. Let $f : [a, b] \to \mathbb{R}^2$ be continuous and be differentiable on (a, b). If $f(a) = f(b)$, then there exists some $c \in (a, b)$ such that $f'(c) = 0$.