DIFFERENTIATION

TSOGTGEREL GANTUMUR

ABSTRACT. We discuss differentiation in $\mathbb{R}^n$ and the inverse function theorem.

CONTENTS

1. CONTINUITY OF SCALAR FUNCTIONS

Let us recall first a definition of continuous functions. Intuitively, a continuous function f sends nearby points to nearby points, i.e, if x is close to y then $f(x)$ is close to $f(y)$.

Definition 1.1. Let $K \subset \mathbb{R}$ be a set. A function $f : K \to \mathbb{R}$ is called *continuous at* $y \in K$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $x \in (y - \delta, y + \delta) \cap K$ implies $|f(x) - f(y)| < \varepsilon$.

In order for f to be continuous at y, first, the value $f(x)$ must be getting closer and closer to some number, say $\alpha \in \mathbb{R}$, as x tends to y, and second, that number α must be equal to the value $f(y)$. The first requirement alone leads to the notion of the *limit of a function*.

Definition 1.2. Let $K \subset \mathbb{R}$ be a set, and let $f : K \to \mathbb{R}$ be a function. We say that $f(x)$ *converges to* $\alpha \in \mathbb{R}$ *as* $x \to y \in \mathbb{R}$, and write

$$f(x) \to \alpha \qquad \text{as} \quad x \to y, \qquad \text{or} \qquad \lim_{x\to y} f(x) = \alpha, \tag{1}$$

if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - \alpha| < \varepsilon$ whenever $0 < |x - y| < \delta$ and $x \in K$. One can write $\lim\limits_{x\in K,\, x\to y} f(x)$, $K \ni x \to y$, etc., to explicitly indicate the domain K.

Remark 1.3. Note that the point y is not required to be in K, and even if $y \in K$, the existence and the value of the limit $\lim\limits_{x\to y} f(x)$ does *not* depend on the value $f(y)$, since we never consider $x = y$ due to the condition $0 < |x - y|$. In other words, we can replace K by $K \setminus \{y\}$ with no effect on the existence and the value of the limit.

Example 1.4. Let $K = [0, 1)$, and let $f : K \to \mathbb{R}$ be a function. Take $y = 2$, and $\alpha \in \mathbb{R}$. Then as long as $\delta \leq 1$, there is *no* x satisfying $0 < |x - y| < \delta$ and $x \in K$. Hence by convention, when $\delta \leq 1$, we must assume that the implication $x \in K$ and $0 < |x-y| < \delta \Rightarrow |f(x)-\alpha| < \varepsilon$

Date: 2016/10/29.

is true for any $\varepsilon > 0$, because there is no x for which we need to check the condition. This means that by Definition 1.2, any possible function $f : [0, 1) \to \mathbb{R}$ would have a limit as $x \to 2$, and this limit can be an arbitrary number $\alpha \in \mathbb{R}$. We could have removed this inconvenience by modifying our definition to require that the set $\{x \in K : 0 < |x - y| < \delta\}$ is nonempty for any $\delta > 0$, but we have not done so because any consideration of situations where y is "far away" from K as in this example would only lead to useless and trivial statements.

The following lemma expresses the limit of a function in terms of the limit of a sequence.

Lemma 1.5. *Let $K \subset \mathbb{R}$ be a set, and let $f : K \to \mathbb{R}$ be a function. Let $y \in \mathbb{R}$ and $\alpha \in \mathbb{R}$. Then $f(x) \to \alpha$ as $x \to y$ if and only if $f(x_n) \to \alpha$ as $n \to \infty$ for every sequence $\{x_n\} \subset K \setminus \{y\}$ converging to y.*

Proof. Suppose that $f(x) \to \alpha$ as $x \to y$, and let $\{x_n\} \subset K \setminus \{y\}$ be a sequence converging to y. We want to show that $f(x_n) \to \alpha$ as $n \to \infty$. Let $\varepsilon > 0$ be arbitrary. Then by definition, there exists $\delta > 0$ such that $0 < |x - y| < \delta$ and $x \in K$ imply $|f(x) - \alpha| < \varepsilon$. Since $x_n \to y$ as $n \to \infty$, there is N such that $|x_n - y| < \delta$ whenever $n > N$. Hence we have $|f(x_n) - \alpha| < \varepsilon$ whenever $n > N$. As $\varepsilon > 0$ is arbitrary, we conclude that $f(x_n) \to \alpha$ as $n \to \infty$.

To prove the other direction, assume that $f(x)$ does *not* converge to α as $x \to y$, i.e., that there is some $\varepsilon > 0$, such that for any $\delta > 0$, there exists some $x \in K$ with $0 < |x-y| < \delta$ and $|f(x) - \alpha| \geq \varepsilon$. In particular, taking $\delta = \frac{1}{n}$, we infer the existence of a sequence $\{x_n\} \subset K$ satisfying $0 < |x_n - y| < \frac{1}{n}$, with $|f(x_n) - \alpha| \geq \varepsilon$ for all n. Thus we have a sequence $\{x_n\} \subset K \setminus \{y\}$ converging to y, with $f(x_n) \not\to \alpha$ as $n \to \infty$. □

An immediate corollary is that continuous functions are precisely the ones that send convergent sequences to convergent sequences. This is sometimes called the *sequential criterion of continuity.*

Corollary 1.6. *Let $K \subset \mathbb{R}$ be a set. Then $f : K \to \mathbb{R}$ is continuous at $y \in K$ if and only if $f(x_n) \to f(y)$ as $n \to \infty$ for every sequence $\{x_n\} \subset K \setminus \{y\}$ converging to x.*

Proof. The second condition is equivalent to $f(x) \to f(y)$ as $x \to y$ by Lemma 1.5, and hence we only need to show that continuity at y is equivalent to $f(x) \to f(y)$ as $x \to y$. Let us explicitly write the definitions of these two concepts side by side to compare.

- Continuity of f at y: For any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$ and $x \in K$.
- Convergence of $f(x)$ to $f(y)$ as $x \to y$: For any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $0 < |x - y| < \delta$ and $x \in K$.

We see that the only difference is in whether we allow $x = y$. Hence it is immediate that continuity implies the convergence property. Now if the convergence property is satisfied, then everything for continuity is there, except the condition $|f(x) - f(y)| < \varepsilon$ when $x = y$. But this is trivially true, because $|f(y) - f(y)| = 0$. □

Exercise 1.7. Let $K \subset \mathbb{R}$ be a set. Show that $f : K \to \mathbb{R}$ is continuous at $y \in K$ if and only if $f(x_n) \to f(y)$ as $n \to \infty$ for every sequence $\{x_n\} \subset K$ converging to y.

Example 1.8. (a) Let $f : \mathbb{R} \to \mathbb{R}$ be the function given by $f(x) = x^2$ for $x \in \mathbb{R}$. Then f is continuous at every point $y \in \mathbb{R}$, because given any sequence $\{x_n\} \subset \mathbb{R}$ converging to y, we have $f(x_n) = x_n^2 \to y^2 = f(y)$ as $n \to \infty$.

(b) Let $g : \mathbb{R} \to \mathbb{R}$ be the function given by $g(x) = |x|$ for $x \in \mathbb{R}$. Then f is continuous at every point $y \in \mathbb{R}$, because given any sequence $\{x_n\} \subset \mathbb{R}$ converging to y, we have $f(x_n) = |x_n| \to |y| = f(y)$ as $n \to \infty$.

(c) We define the *Heaviside step function* $\theta : \mathbb{R} \to \mathbb{R}$ by

$$\theta(x) = \begin{cases} 1 & \text{for } x > 0, \\ 0 & \text{for } x \le 0. \end{cases} \tag{2}$$

It is clear that θ is continuous at every $x \in \mathbb{R} \setminus \{0\}$. Our intuition tells us that θ is not continuous at $x = 0$. Indeed, let $x_n = \frac{1}{n}$ and $y_n = -\frac{1}{n}$ for $n \in \mathbb{N}$. Then we have $x_n \to 0$ and $y_n \to 0$, but $\theta(x_n) \to 1$ and $\theta(y_n) \to 0$ as $n \to \infty$. Since $1 \neq 0$, the sequential criterion of continuity implies that θ is not continuous at $x = 0$.

(d) The *Dirichlet function* $h : \mathbb{R} \to \mathbb{R}$ is defined by

$$h(x) = \begin{cases} 1 & \text{for } x \in \mathbb{Q}, \\ 0 & \text{for } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases} \tag{3}$$

For any $x \in \mathbb{R}$, we can find two sequences $\{x_n\} \subset \mathbb{Q}$ and $\{y_n\} \subset \mathbb{R} \setminus \mathbb{Q}$ satisfying $x_n \to x$ and $y_n \to x$ as $n \to \infty$. Since $h(x_n) = 1$ and $h(y_n) = 0$, we have $h(x_n) \to 1$ and $h(y_n) \to 0$, and hence we conclude that h is not continuous at any point $x \in \mathbb{R}$.

Exercise 1.9. In each of the following cases, verify if the value $f(0)$ can be defined so that the resulting function f is continuous in $\mathbb{R}$. Choose from the following phrases to best describe the situation in each case: jump discontinuity, removable singularity, blow up or pole, essential/oscillatory singularity.

(a) $f(x) = \dfrac{1}{x}$ (b) $f(x) = \dfrac{1}{|x|}$ (c) $f(x) = \dfrac{|x|}{x}$ (d) $f(x) = \dfrac{\sin x}{x}$

(e) $f(x) = \sin \frac{1}{x}$ (f) $f(x) = x \sin \frac{1}{x}$ (g) $f(x) = \frac{1}{x} \sin \frac{1}{x}$

Now we want to introduce a way to compare asymptotic magnitudes of two functions.

Definition 1.10 (Little 'o' notation)**.** Let $K \subset \mathbb{R}$ be a set, let $y \in \mathbb{R}$, and let $f : K \to \mathbb{R}$ and $g : K \to \mathbb{R}$ be functions. Then we write

$$f(x) = o(g(x)) \qquad \text{as} \quad x \to y, \tag{4}$$

to mean that

$$\frac{f(x)}{g(x)} \to 0 \qquad \text{as} \quad x \to y. \tag{5}$$

Furthermore, for $h : K \to \mathbb{R}$, the notation

$$f(x) = h(x) + o(g(x)) \qquad \text{as} \quad x \to y, \tag{6}$$

is understood to be

$$f(x) - h(x) = o(g(x)) \qquad \text{as} \quad x \to y. \tag{7}$$

Example 1.11. (a) We have $x^3 = o(x^2)$ as $x \to 0$, because $\frac{x^3}{x^2} \to 0$ as $x \to 0$.

(b) We have $\sin x \neq o(x)$ as $x \to 0$, because $\frac{\sin x}{x} \not\to 0$ as $x \to 0$.

(c) We have $\sqrt{x} = \sqrt{2} + o(1)$ as $x \to 2$, because $\frac{\sqrt{x}-\sqrt{2}}{1} \to 0$ as $x \to 2$.

(d) We have $(1+x)^2 = 1 + 2x + o(x)$ as $x \to 0$, because $\frac{(1+x)^2-1-2x}{x} \to 0$ as $x \to 0$.

Exercise 1.12. Verify if the following statements are true.

(a) $\sin x = 1 + o(x - \frac{\pi}{2})$ as $x \to \frac{\pi}{2}$.
(b) $\cos x = \sin x + o(1)$ as $x \to 0$.
(c) $\tan x = \cot x + o(1)$ as $x \to \frac{\pi}{4}$.
(d) $\log x = o(x)$ as $x \to 0$.
(e) $1 = o(\tan x)$ as $x \to \frac{\pi}{2}$.

Remark 1.13. We see that the following statements are equivalent.

- $f(x) \to \alpha$ as $x \to y$.
- $x_n \to y$ implies $f(x_n) \to \alpha$, as $n \to \infty$.
- $f(x) = \alpha + o(1)$ as $x \to y$.

Similarly, the following statements are equivalent.

- f is continuous at y.
- $f(x) \to f(y)$ as $x \to y$.
- $x_n \to y$ implies $f(x_n) \to f(y)$, as $n \to \infty$.
- $f(x) = f(y) + o(1)$ as $x \to y$.

Remark 1.14. Intuitively, the condition $f(x) = f(y) + o(1)$ says that if f is continuous at y, the value $f(x)$ can be approximated by the constant $f(y)$ with the error of $o(1)$.

2. Differentiability of scalar functions

Let us recall the usual definition of differentiability. This is essentially the definition introduced by Augustin-Louis Cauchy in 1821.

Definition 2.1. Let $K \subset \mathbb{R}$ be a set, and let $f : K \to \mathbb{R}$ be a function. We say that f is *differentiable at* $y \in K$, if there exists $\lambda \in \mathbb{R}$ such that

$$\frac{f(x) - f(y)}{x - y} \to \lambda \qquad \text{as} \quad K \ni x \to y. \tag{8}$$

We call $f'(y) = \lambda$ the *derivative of* f *at* y. If f is differentiable at each point of K, then f is said to be *differentiable in* K.

We now prove several useful criteria of differentiability. In the following lemma, (b) is called the *sequential criterion*, (c) is the criterion introduced by Constantin Carathéodory in 1950, and finally, (d) is introduced by Karl Weierstrass in his 1861 lectures.

Lemma 2.2. *Let $K \subset \mathbb{R}$, let $y \in K$, and let $f : K \to \mathbb{R}$ be a function. Then the following are equivalent.*

(a) f is differentiable at y.

(b) There exists a number $\lambda \in \mathbb{R}$, such that

$$\frac{f(x_n) - f(y)}{x_n - y} \to \lambda \qquad \textit{as} \quad n \to \infty, \tag{9}$$

for every sequence $\{x_n\} \subset K \setminus \{y\}$ converging to y.

(c) There exists a function $g : K \to \mathbb{R}$, continuous at y, such that

$$f(x) = f(y) + g(x)(x - y) \qquad \textit{for} \quad x \in K. \tag{10}$$

(d) There exists a number $\lambda \in \mathbb{R}$, such that

$$f(x) = f(y) + \lambda(x - y) + o(x - y) \qquad \textit{as} \quad K \ni x \to y. \tag{11}$$

Proof. Equivalence of (a) and (b) is immediate from Lemma 1.5.

Suppose that (a) holds. We define the function $g : K \to \mathbb{R}$ by

$$g(x) = \begin{cases} \frac{f(x)-f(y)}{x-y} & \text{for} \quad x \in K \setminus \{y\} \\ f'(y) & \text{for} \quad y = x. \end{cases} \tag{12}$$

This function satisfies (10) by construction. Since $g(x) \to f'(y) \equiv g(y)$ as $x \to y$ by differentiability, g is continuous at y. Hence (c) holds.

Now suppose that (c) holds. Let $\lambda = g(y)$ and $h(x) = g(x) - \lambda$. Then h is continuous at y with $h(y) = 0$, and

$$f(x) = f(y) + \lambda(x - y) + h(x)(x - y) \qquad \text{for} \quad x \in K. \tag{13}$$

This implies (d), since

$$\frac{h(x)(x-y)}{x-y} = h(x) \to 0 \qquad \text{as} \quad K \ni x \to y, \tag{14}$$

by continuity of h and the fact that $h(y) = 0$.

Finally, let (d) hold. By definition of the little 'o' notation (Definition 1.10), this means that

$$\frac{f(x) - f(y) - \lambda(x-y)}{x-y} \to 0 \qquad \text{as} \quad K \ni x \to y, \tag{15}$$

or equivalently,

$$\frac{f(x) - f(y)}{x-y} - \lambda \to 0 \qquad \text{as} \quad K \ni x \to y. \tag{16}$$

Thus (a) holds, i.e., f is differentiable at y, cf. Definition 2.1. □

Remark 2.3. We see from (11) that if f is differentiable at y then f is continuous at y, since $\lambda(x-y) + o(x-y) = o(1)$ as $x \to y$.

Remark 2.4. By (11), differentiability of f at y is equivalent to the condition that $f(x)$ can be approximated by the linear function $\ell(x) = f(y) + \lambda(x-y)$ with the error of $o(x-y)$. Recall that continuity of f at y is equivalent to saying that $f(x)$ can be approximated by the constant $f(y)$ with the error of $o(1)$, cf. Remark 1.14.

Example 2.5. (a) Let $c \in \mathbb{R}$, and let $f(x) = c$ be a constant function. Then since $f(x) = f(y) + 0 \cdot (x-y)$ for all x, y, we get $f'(y) = 0$ for all y.

(b) Let $a, c \in \mathbb{R}$, and let $f(x) = ax + c$ be a linear (also known as affine) function. Since $f(x) = f(y) + a(x-y)$ for all x, y, we get $f'(y) = a$ for all y.

(c) Let $f(x) = \frac{1}{x}$, fix $y \in \mathbb{R} \setminus \{0\}$, and for $x \in \mathbb{R} \setminus \{0, y\}$ define

$$g(x) = \frac{\frac{1}{x} - \frac{1}{y}}{x-y} = -\frac{1}{xy}. \tag{17}$$

Upon defining $g(y) = -\frac{1}{y^2}$, the function $g(x) = -\frac{1}{x} \cdot \frac{1}{y}$ becomes continuous at $x = y$, and therefore f is differentiable at y with

$$f'(y) = \big(\frac{1}{y}\big)' = -\frac{1}{y^2} \qquad (y \neq 0). \tag{18}$$

(d) Let us try to differentiate $f(x) = |x|$ at $x = 0$. With $x_n = \frac{1}{n}$ for $n \in \mathbb{N}$, we have $\{x_n\} \subset \mathbb{R} \setminus \{0\}$ and $x_n \to 0$ as $n \to \infty$. On one hand, we get

$$\lim_{n\to\infty} \frac{f(x_n) - f(0)}{x_n - 0} = \lim_{n\to\infty} \frac{|x_n|}{x_n} = 1, \tag{19}$$

but on the other hand, with $y_n = -x_n$, we infer

$$\lim_{n\to\infty} \frac{f(y_n) - f(0)}{y_n - 0} = \lim_{n\to\infty} \frac{|y_n|}{y_n} = -\lim_{n\to\infty} \frac{x_n}{x_n} = -1. \tag{20}$$

The definition of derivative requires these two limits to be the same, and thus we conclude that $f(x) = |x|$ is *not* differentiable at $x = 0$.

(e) Consider the differentiability of $f(x) = \sqrt[3]{x}$ at $x = 0$. Let $x_n = \frac{1}{n^3}$. It is obvious that $x_n \neq 0$ and $x_n \to 0$. We have

$$\frac{f(x_n) - f(0)}{x_n - 0} = \frac{\sqrt[3]{x_n}}{x_n} = n^2, \tag{21}$$

which diverges as $n \to \infty$. Hence $f(x) = \sqrt[3]{x}$ is *not* differentiable at $x = 0$.

Exercise 2.6. Show that $f(x) = x^n$ is differentiable in $\mathbb{R}$, for $n \in \mathbb{N}$, with $f'(x) = nx^{n-1}$.

Exercise 2.7. Let $K \subset K$, and suppose that $f : K \to \mathbb{R}$ satisfies

$$f(x) = a + bx + o(x - y) \qquad \text{as} \quad x \to y \in K, \tag{22}$$

for some constants $a, b \in \mathbb{R}$. Show that f is differentiable at y with $f'(y) = b$, and $f(y) = a+by$.

We now review differentiability of various combinations of differentiable functions.

Theorem 2.8. *Let $f, g : (a, b) \to \mathbb{R}$ be functions differentiable at $x \in (a, b)$. Then the following are true.*

a) The sum and difference $f \pm g$ are differentiable at x, with

$$(f \pm g)'(x) = f'(x) \pm g'(x). \tag{23}$$

These are called the sum and difference rules*.*

b) The product fg is differentiable at x, with

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x). \tag{24}$$

This is called the product rule*.*

c) If $F : (c, d) \to \mathbb{R}$ is a function differentiable at $g(x)$, with $g((a, b)) \subset (c, d)$, then the composition $F \circ g : (a, b) \to \mathbb{R}$ is differentiable at x, with

$$(F \circ g)'(x) = F'(g(x))g'(x). \tag{25}$$

This is called the chain rule*.*

d) If $f : (a, b) \to f((a, b))$ is bijective and $f'(x) \neq 0$, then the inverse $f^{-1} : f((a, b)) \to (a, b)$ is differentiable at $y = f(x)$, with

$$(f^{-1})'(y) = \frac{1}{f'(x)}. \tag{26}$$

Proof. b) By definition, there is a function $\tilde{f} : (a, b) \to \mathbb{R}$, continuous at x, satisfying

$$f(y) = f(x) + \tilde{f}(y)(y - x), \qquad y \in (a, b), \tag{27}$$

and $f'(x) = \tilde{f}(x)$. Similarly, there is a function $\tilde{g} : (a, b) \to \mathbb{R}$, continuous at x, and with $g'(x) = \tilde{g}(x)$, such that

$$g(y) = g(x) + \tilde{g}(y)(y - x), \qquad y \in (a, b). \tag{28}$$

By multiplying (27) and (28), we get

$$\begin{aligned} f(y)g(y) &= f(x)g(x) + g(x)\tilde{f}(y)(y - x) + f(x)\tilde{g}(y)(y - x) + \tilde{f}(y)\tilde{g}(y)(y - x)^2 \\ &= f(x)g(x) + [g(x)\tilde{f}(y) + f(x)\tilde{g}(y) + \tilde{f}(y)\tilde{g}(y)(y - x)](y - x). \end{aligned} \tag{29}$$

The expression in the square brackets, as a function of y, is continuous at $y = x$, with

$$\begin{aligned} [g(x)\tilde{f}(y) + f(x)\tilde{g}(y) + \tilde{f}(y)\tilde{g}(y)(y - x)]\big|_{y=x} &= g(x)\tilde{f}(x) + f(x)\tilde{g}(x) \\ &= g(x)f'(x) + f(x)g'(x), \end{aligned} \tag{30}$$

which shows that fg is differentiable at x, and that (24) holds.

c) Since F is differentiable at $g(x)$, by definition, there is a function $\tilde{F} : (c, d) \to \mathbb{R}$, continuous at $g(x)$, and with $F'(g(x)) = \tilde{F}(g(x))$, such that

$$F(z) = F(g(x)) + \tilde{F}(z)(z - g(x)), \qquad z \in (c, d). \tag{31}$$

Plugging $z = g(y)$ into (31), we get

$$F(g(y)) = F(g(x)) + \tilde{F}(g(y))(g(y) - g(x)) = F(g(x)) + \tilde{F}(g(y))\tilde{g}(y)(y - x), \tag{32}$$

where in the last step we have used (28). The function $y \mapsto \tilde{F}(g(y))\tilde{g}(y)$ is continuous at $y = x$, with $\tilde{F}(g(x))\tilde{g}(x) = F'(g(x))g'(x)$, which confirms that $F \circ g$ is differentiable at x, and that (25) holds.

d) By definition, there is $g : (a, b) \to \mathbb{R}$, continuous at x, with $g(x) = f'(x) \neq 0$, such that

$$f(z) = f(x) + g(z)(z - x) \qquad \text{for } z \in (a, b). \tag{33}$$

Since g is continuous at x, we infer the existence if an open interval $(c, d) \ni x$ such that $g(z) \neq 0$ for all $z \in (c, d)$. For $t \in f((c, d))$, we have $z = f^{-1}(t) \in (c, d)$, and

$$f^{-1}(t) - f^{-1}(y) = z - x = \frac{f(z) - f(x)}{g(z)} = \frac{t - y}{g(f^{-1}(t))}. \tag{34}$$

The function $\frac{1}{g(f^{-1}(t))}$ is continuous at $t = y$, meaning that f^{-1} is differentiable at y, and that (26) holds. □

Exercise 2.9. Prove a) of the preceding theorem.

Exercise 2.10. Prove that differentiability is a local property, in the sense that $f : (a, b) \to \mathbb{R}$ is differentiable at $y \in (a, b)$ if and only if $g = f|_{(y-\varepsilon,y+\varepsilon)}$ is differentiable at y, where $\varepsilon > 0$ is small. Here $g = f|_{(y-\varepsilon,y+\varepsilon)}$ means that $g : (y - \varepsilon, y + \varepsilon) \to \mathbb{R}$ is defined by $g(x) = f(x)$ for $x \in (y - \varepsilon, y + \varepsilon)$. We say that g is the *restriction* of f to the interval $(y - \varepsilon, y + \varepsilon)$.

Example 2.11. (a) By the product rule, we have

$$\begin{aligned} &(x^2)' = 1 \cdot x + x \cdot 1 = 2x, \\ &(x^3)' = (x^2 \cdot x)' = 2x \cdot x + x^2 \cdot 1 = 3x^2, \ldots \\ &(x^n)' = nx^{n-1} \qquad (n \in \mathbb{N}). \end{aligned} \tag{35}$$

(b) By the sum and product rules, all polynomials are differentiable in $\mathbb{R}$, and the derivative of a polynomial is again a polynomial.

(c) Given a function $f : (a, b) \to \mathbb{R}$ that does not vanish anywhere in (a, b), we can write the reciprocal function $\frac{1}{f}$ as $F \circ f$ with $F(z) = \frac{1}{z}$. If f is differentiable at $x \in (a, b)$, then by the chain rule, $\frac{1}{f}$ is differentiable at x and

$$(\frac{1}{f})'(x) = (F \circ f)'(x) = F'(f(x))f'(x) = -\frac{f'(x)}{[f(x)]^2}. \tag{36}$$

In particular, we have

$$(x^{-n})' = -\frac{nx^{n-1}}{x^{2n}} = -nx^{-n-1} \qquad (n \in \mathbb{N}). \tag{37}$$

(d) Let $f(x) = x^n$ for $x \in [0, \infty)$, where $n \in \mathbb{N}$. We have $f'(x) = nx^{n-1}$ at $x > 0$, and the inverse function is the arithmetic n-th root $f^{-1}(y) = \sqrt[n]{y}$ $(y \geq 0)$. Since $f'(x) > 0$ for $x > 0$, the inverse f^{-1} is differentiable at each $y > 0$, with

$$(f^{-1})'(y) = \frac{1}{f'(f^{-1}(y))} = \frac{1}{n(\sqrt[n]{y})^{n-1}} = \frac{1}{n} y^{\frac{1-n}{n}}. \tag{38}$$

Moreover, by the chain rule, for $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, we infer

$$(x^{\frac{m}{n}})' = ((\sqrt[n]{x})^m)' = m(\sqrt[n]{x})^{m-1} \cdot \frac{1}{n} x^{\frac{1-n}{n}} = \frac{m}{n} x^{\frac{m-1}{n} + \frac{1-n}{n}} = \frac{m}{n} x^{\frac{m}{n} - 1}, \tag{39}$$

that is

$$(x^a)' = ax^{a-1} \qquad \text{at each } x > 0, \quad \text{for } a \in \mathbb{Q}. \tag{40}$$

Exercise 2.12. Let $f, g : (a, b) \to \mathbb{R}$ be functions differentiable at $x \in (a, b)$, with $g(x) \neq 0$. Show that the quotient f/g is differentiable at x, and the following *quotient rule* holds.

$$(\frac{f}{g})'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}. \tag{41}$$

Compute the derivative of $q(x) = \frac{3x^3}{x^2+1}$.

3. Continuity of vector functions

The set of all *ordered pairs* of real numbers is denoted by

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\}. \tag{42}$$

Ordered means that, for instance, $(1, 3) \neq (3, 1)$. As an example, the position of a point on the surface of the Earth can be described by an element of $\mathbb{R}^2$, by its latitude and longitude. More generally, for $n \in \mathbb{N}$, we let

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ times}} = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\}. \tag{43}$$

An element of $\mathbb{R}^n$ is called an n-tuple, an n-vector, a point, or simply a *vector*. In a context where both $\mathbb{R}$ and $\mathbb{R}^n$ (with $n > 1$) are present, an element of $\mathbb{R}$ (i.e., a real number) is called a *scalar*. Given a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the number $x_k \in \mathbb{R}$ is called the *k-th component of x*, for $k \in \{1, \ldots, n\}$.

Example 3.1. Consider n foreign currencies, and let x_k be the exchange rate between the k-th currency and Canadian dollar (at a certain moment of time). Then any possible outcome $(x_1, \ldots, x_n)$ can be considered as an element of $\mathbb{R}^n$.

Let $K \subset \mathbb{R}$, and let $f : K \to \mathbb{R}^n$ be a function. Such functions are called *vector valued functions* (of a single variable), or *vector functions*. In contrast, $\mathbb{R}$-valued functions (i.e., $n = 1$) are called *scalar valued functions*, or *scalar functions*. For $t \in K$, the value $f(t)$ is an n-vector; Let us denote the k-th component of $f(t) \in \mathbb{R}^n$ by $f_k(t) \in \mathbb{R}$. Since t can be any point in K, this defines a function $f_k : K \to \mathbb{R}$, called the *k-th component of f*, for each $k \in \{1, \ldots, n\}$. Thus a vector valued function is simply a collection of scalar valued functions.

In this and the next sections, we will study vector valued functions of a single variable. It is more or less straightforward, but will give us a chance to introduce notations and concepts that will be important later.

First, let us briefly discuss vector sequences. A *vector sequence* is simply a sequence $x^{(1)}, x^{(2)}, \ldots, x^{(i)}, \ldots$, consisting of vectors $x^{(i)} \in \mathbb{R}^n$, $i \in \mathbb{N}$. It can also be thought of as a function $f : \mathbb{N} \to \mathbb{R}^n$, whose domain is the set of natural numbers. The correspondence can be defined by identifying the i-th term $x^{(i)} \in \mathbb{R}^n$ of the sequence with the value $f(i) \in \mathbb{R}^n$ of the function at $i \in \mathbb{N}$. The k-th component of $x^{(i)} \in \mathbb{R}^n$ is denoted by $x_k^{(i)} \in \mathbb{R}$, for $k \in \{1, \ldots, n\}$. If we fix $k \in \{1, \ldots, n\}$, then $\{x_k^{(1)}, x_k^{(2)}, \ldots\}$ is a scalar (i.e., real number) sequence, called the *k-th component* of the vector sequence $\{x^{(i)}\}$. Hence a vector sequence is simply a collection of scalar sequences.

Definition 3.2. We say that a vector sequence $\{x^{(i)}\} \subset \mathbb{R}^n$ *converges to* $y \in \mathbb{R}^n$ if for each $k \in \{1, \ldots, n\}$, the k-th component of $\{x^{(i)}\}$ converges to the k-th component of y. That is, we write $x^{(i)} \to y$ as $i \to \infty$ if $x_k^{(i)} \to y_k$ as $i \to \infty$ for each $k \in \{1, \ldots, n\}$.

Remark 3.3. Let us write out the preceding definition explicitly. Thus $x^{(i)} \to y$ as $i \to \infty$ if and only if for each $k \in \{1, \ldots, n\}$, and for any $\varepsilon > 0$, there exists an index $N_{k,\varepsilon}$ such that $|x_k^{(i)} - y_k| < \varepsilon$ whenever $i > N_{k,\varepsilon}$. In this setting, let $N_\varepsilon = \max\{N_{1,\varepsilon}, \ldots, N_{n,\varepsilon}\}$ for $\varepsilon > 0$. Then we have $|x_k^{(i)} - y_k| < \varepsilon$ whenever $i > N_\varepsilon$ for each k and for any $\varepsilon > 0$. In other words, if $x^{(i)} \to y$ as $i \to \infty$, then for any $\varepsilon > 0$, there exists an index N_ε such that $\max\limits_{k=1,\ldots,n} |x_k^{(i)} - y_k| < \varepsilon$ whenever $i > N_\varepsilon$. It is obvious that if the latter condition holds, then $x^{(i)} \to y$ as $i \to \infty$. We have expressed the convergence $x^{(i)} \to y$ of a vector sequence in terms of the convergence $d_i \to 0$ of the scalar sequence $\{d_i\}$ defined by $d_i = \max\limits_{k=1,\ldots,n} |x_k^{(i)} - y_k|$. We can think of d_i as some kind of "distance" between the points $x^{(i)}$ and y, so that the convergence $x^{(i)} \to y$ is identified with the convergence of the distance $d_i \to 0$.

Definition 3.4. We define the *maximum norm* of $x \in \mathbb{R}^n$ as

$$|x|_\infty = \max_{k=1,\ldots,n} |x_k|. \tag{44}$$

Moreover, for $x, y \in \mathbb{R}^n$ and $t \in \mathbb{R}$, we define

$$x \pm y = (x_1 \pm y_1, \ldots, x_n \pm y_n) \qquad \text{and} \qquad tx = xt = (tx_1, \ldots, tx_n). \tag{45}$$

Example 3.5. For $x = (1, -3) \in \mathbb{R}^2$, we have $|x|_\infty = 3$. Moreover, $2x = x \cdot 2 = (2, -6)$, and $x + (-2, 3) = (-1, 0)$.

In light of the new notations, Remark 3.3 says that the convergence $x^{(i)} \to y$ is equivalent to the scalar sequence $\{d_i\}$ defined by $d_i = |x^{(i)} - y|_\infty$ being convergent to 0. Let us summarize it in the following lemma.

Lemma 3.6. *A sequence $\{x^{(i)}\} \subset \mathbb{R}^n$ converges to $y \in \mathbb{R}^n$ iff $|x^{(i)} - y|_\infty \to 0$ as $i \to \infty$.*

Analogously to the limit of a vector sequence, we initially define the limit of a vector function component-wise, and then express it in terms of the maximum norm.

Definition 3.7. Let $K \subset \mathbb{R}$ be a set, and let $f : K \to \mathbb{R}^n$ be a vector function. We say that $f(x)$ *converges to* $\alpha \in \mathbb{R}^n$ *as* $x \to y \in \mathbb{R}$, and write

$$f(x) \to \alpha \qquad \text{as} \quad x \to y, \qquad \text{or} \qquad \lim_{x \to y} f(x) = \alpha, \tag{46}$$

if $f_k(x) \to \alpha_k$ as $x \to y$, for each $k \in \{1, \ldots, n\}$.

Exercise 3.8. In the context of this definition, show that the following are equivalent.

- $f(x) \to \alpha$ as $x \to y$.
- $|f(x) - \alpha|_\infty \to 0$ as $x \to y$.
- $f(x_i) \to \alpha$ as $i \to \infty$ for every sequence $\{x_i\} \subset K \setminus \{y\}$ converging to y.

Finally, we are ready to define continuity for vector functions of a single variable.

Definition 3.9. Let $K \subset \mathbb{R}$ be a set, and let $f : K \to \mathbb{R}^n$ be a vector function. We say that f is *continuous at* $y \in K$ if each component of f is continuous at y, or equivalently, if $f(x) \to f(y)$ as $x \to y$.

Example 3.10. (a) The function $f : \mathbb{R} \to \mathbb{R}^2$ defined by $f(x) = (x \cos x, x \sin x)$ is obviously continuous at each $x \in \mathbb{R}$.

(b) The function $f(x) = (\theta(x), x^2)$ is discontinuous at 0, and continuous everywhere in $\mathbb{R} \setminus \{0\}$.

Exercise 3.11. In the context of this definition, show that the following are equivalent.

- f is continuous at $y \in K$.
- $|f(x) - f(y)|_\infty \to 0$ as $x \to y$.
- The scalar function $d : K \to \mathbb{R}$ defined by $d(x) = |f(x) - f(y)|_\infty$ is continuous at y, with $d(y) = 0$.
- $f(x_i) \to f(y)$ as $i \to \infty$ for every sequence $\{x_i\} \subset K \setminus \{y\}$ converging to y.

In particular, from the preceding exercise, we see that continuity of f at y is equivalent to $f(x) = f(y) + e(x)$, with $|e(x)|_\infty \to 0$ as $x \to y$, that is, $f(x)$ is approximated by the constant vector $f(y)$ with the error of $o(1)$, when the error is measured in the maximum norm. To be able to write this fact in one line we extend the little 'o' notation to vector valued functions.

Definition 3.12 (Little 'o' notation)**.** Let $K \subset \mathbb{R}$ be a set, let $y \in \mathbb{R}$, and let $f : K \to \mathbb{R}^n$ and $g : K \to \mathbb{R}$. Then we write

$$f(x) = o(g(x)) \qquad \text{as} \quad x \to y, \tag{47}$$

to mean that

$$\frac{|f(x)|_\infty}{|g(x)|} \to 0 \qquad \text{as} \quad x \to y. \tag{48}$$

Furthermore, for $h : K \to \mathbb{R}^n$, the notation

$$f(x) = h(x) + o(g(x)) \qquad \text{as} \quad x \to y, \tag{49}$$

is understood to be

$$f(x) - h(x) = o(g(x)) \qquad \text{as} \quad x \to y. \tag{50}$$

Remark 3.13. Continuity of f at y is equivalent to $f(x) = f(y) + o(1)$ as $x \to y$.

4. Differentiability of vector functions

Similarly to continuity, differentiability of vector functions is defined component-wise.

Definition 4.1. Let $K \subset \mathbb{R}$ be a set, and let $f : K \to \mathbb{R}^n$ be a vector function. We say that f is *differentiable at* $y \in K$, if each component of f is differentiable at y. We call $f'(y) = (f_1'(y), \ldots, f_n'(y)) \in \mathbb{R}^n$ the *derivative of* f *at* y. If f is differentiable at each point of K, then f is said to be *differentiable in* K.

Lemma 4.2. *Let $K \subset \mathbb{R}$, let $y \in K$, and let $f : K \to \mathbb{R}^n$ be a vector function. Then the following are equivalent.*

(a) f is differentiable at y.

(b) There exists a function $g : K \to \mathbb{R}^n$, continuous at y, such that

$$f(x) = f(y) + g(x)(x - y) \qquad \textit{for} \quad x \in K. \tag{51}$$

(c) There exists a vector $\lambda \in \mathbb{R}^n$, such that

$$f(x) = f(y) + \lambda(x - y) + o(x - y) \qquad \textit{as} \quad K \ni x \to y. \tag{52}$$

(d) There exists a vector $\lambda \in \mathbb{R}^n$, such that

$$\frac{f(x_i) - f(y)}{x_i - y} \to \lambda \qquad \textit{as} \quad i \to \infty, \tag{53}$$

for every sequence $\{x_i\} \subset K \setminus \{y\}$ converging to y.

Proof. Let f be differentiable at y. Then by definition, for each k, there exists $g_k : K \to \mathbb{R}$, continuous at y, such that

$$f_k(x) = f_k(y) + g_k(x)(x - y) \qquad \text{for} \quad x \in K. \tag{54}$$

The vector function $g : K \to \mathbb{R}^n$ defined by $g(x) = (g_1(x), \ldots, g_n(x))$ is clearly continuous at y, and satisfies the condition in (b).

Suppose that (b) holds. Let $\lambda = g(y)$ and let $h(x) = g(x) - \lambda$. Then h is continuous at y with $h(y) = 0$, and we have

$$f(x) = f(y) + \lambda(x - y) + h(x)(x - y) \qquad \text{for} \quad x \in K. \tag{55}$$

This implies (c), since

$$\frac{|h(x)(x - y)|_\infty}{|x - y|} = |h(x)|_\infty \to 0 \qquad \text{as} \quad K \ni x \to y, \tag{56}$$

where we have taken into account the fact that $|ta|_\infty = |t| \cdot |a|_\infty$ for $t \in \mathbb{R}$ and $a \in \mathbb{R}^n$.

Now suppose that (c) holds, that is,

$$\frac{f(x)-f(y)}{x-y}-\lambda=\frac{f(x)-f(y)-\lambda(x-y)}{x-y}\to 0 \qquad \text{as} \quad K\ni x\to y. \tag{57}$$

In view of Exercise 3.8, this means that for every sequence $\{x_i\}\subset K\setminus\{y\}$ converging to y, we have (53), which is (d).

Finally, suppose that (d) holds. In components, it reads as follows. For each k, we have

$$\frac{f_k(x_i)-f_k(y)}{x_i-y}\to\lambda_k \qquad \text{as} \quad i\to\infty, \tag{58}$$

for every sequence $\{x_i\}\subset K\setminus\{y\}$ converging to y. Hence each component of f is differentiable at y, that is, f is differentiable at y, cf. Definition 4.1 □

Remark 4.3. Functions $\ell:\mathbb{R}\to\mathbb{R}^n$ of the form

$$\ell(x)=\alpha+\beta x, \tag{59}$$

where $\alpha,\beta\in\mathbb{R}^n$ are fixed vectors, are called *linear functions*. By (c) of the preceding lemma, f is differentiable at y if and only if $f(x)$ can be approximated by a linear function with the error of $o(x-y)$. The "only if" part is immediate, because $\ell(x)=f(y)+f'(y)(x-y)$ is a linear function. In the other direction, since $\ell(x)\to f(y)$ as $x\to y$, we get $f(y)=\alpha+\beta y$, and thus $\ell(x)=f(y)+\beta(x-y)$.

Exercise 4.4. Let $f:(a,b)\to\mathbb{R}^n$ and $\phi:(a,b)\to\mathbb{R}$ be both differentiable at $y\in(a,b)$. Show that the product $\phi f:(a,b)\to\mathbb{R}^n$ is differentiable at y, with

$$(\phi f)'(y)=\phi'(y)f(y)+\phi(y)f'(y).$$

Exercise 4.5. Let $f:(a,b)\to\mathbb{R}^n$ and $\phi:(c,d)\to(a,b)$, where ϕ is differentiable at $t\in(c,d)$, and f is differentiable at $\phi(t)\in(a,b)$. Show that the composition $f\circ\phi:(c,d)\to\mathbb{R}^n$ is differentiable at t, with

$$(f\circ\phi)'(y)=f'(\phi(t))\phi'(t).$$

5. Functions of several variables: Continuity

In this section, we start our study of functions of several variables. A function of several variables is simply a function $f:K\to\mathbb{R}$ or $f:K\to\mathbb{R}^m$, where $K\subset\mathbb{R}^n$. For now, we will keep $m=1$, that is, we temporarily focus on scalar valued functions. Examples of such functions are given by $f(x)=\log(x_1+x_2)$ with $K=\{x\in\mathbb{R}^2:x_1+x_2>0\}$, and $f(x)=-|x|_\infty$ with $K=\mathbb{R}^n$. If $x=(x_1,x_2,\ldots,x_n)$, we have $f(x)=f((x_1,x_2,\ldots,x_n))$, but we typically omit one set of brackets and simply write it as $f(x)=f(x_1,x_2,\ldots,x_n)$.

The first question is how we define continuity for functions of several variables. For functions of the sort $g:\mathbb{R}\to\mathbb{R}^n$, we defined continuity as continuity of its components $g_k:\mathbb{R}\to\mathbb{R}$, and then reformulated it in terms of the maximum norm, which gives a way to measure the distance between two points in $\mathbb{R}^n$. For a function of the sort $f:\mathbb{R}^n\to\mathbb{R}$, taking $n=2$ for simplicity, the component-wise approach to continuity would be to require that the single variable functions $h_1(t)=f(t,x_2)$ with x_2 fixed, and $h_2(s)=f(x_1,s)$ with x_1 fixed, are both continuous. In this context, the distance-based approach to continuity would be to require that the values $f(x)$ and $f(y)$ be close when the points x and y are close, in the sense that $|x-y|_\infty$ is small. It is interesting to note that Cauchy wrote in his 1821 book that these two approaches would lead to the same notion of continuity, just as in the case of $g:\mathbb{R}\to\mathbb{R}^n$, but it was later discovered that the two notions are different.

Definition 5.1. Let $K\subset\mathbb{R}^n$ be a set, and let $f:K\to\mathbb{R}$ be a function. We say that f is *separately continuous at* $y\in K$, if for each k, the restriction $g_k(t)=f(y_1,\ldots,y_{k-1},t,y_{k+1},\ldots,y_n)$ is continuous at $t=y_k$.

Example 5.2. (a) The function $f(x) = x_1^2 + \sin x_2$ is separately continuous at $(0, \frac{\pi}{2}) \in \mathbb{R}^2$, because both $g_1(t) = t^2 + 1$ and $g_2(t) = \sin t$ are continuous at $t = 0$ and at $t = \frac{\pi}{2}$, respectively.

(b) At times, it is convenient to denote a vector in $\mathbb{R}^2$ by (x, y) with x and y being real numbers, instead of using subscripts as in (x_1, x_2). Then the value of $f : \mathbb{R}^2 \to \mathbb{R}$ at $(x, y) \in \mathbb{R}^2$ is $f(x, y)$. Consider

$$f(x,y) = \begin{cases} 1 & \text{for } x < y < 3x \\ 0 & \text{otherwise.} \end{cases} \tag{60}$$

This function is separately continuous at $(0, 0)$ with $f(x, y) = 0$, because $f(t, 0) = f(0, t) = 0$ for all $t \in \mathbb{R}$. However, there exist points (x, y) that are arbitrarily close to $(0, 0)$ with $f(x, y) = 1$, such as $(x, 2x)$ for $x > 0$.

We see that separate continuity of $f : \mathbb{R}^2 \to \mathbb{R}$, say, at $(0, 0)$, imposes conditions only on the two axes, and hence it is not dependent on the behaviour of f at points such as (x, x) with $x > 0$ arbitrarily small. The following is a stronger definition, which follows the distance-based approach that we have discussed. We use the sequential criterion as the initial definition, and will establish an ε-δ criterion later.

Definition 5.3. Let $K \subset \mathbb{R}^n$ be a set, and let $f : K \to \mathbb{R}$ be a function. We say that f is *jointly continuous* or simply *continuous at* $y \in K$, if $f(x^{(i)}) \to f(y)$ as $i \to \infty$ for every sequence $\{x^{(i)}\} \setminus \{y\} \subset K$ converging to y.

Example 5.4. In Example 5.2(b), we have $f(\frac{1}{m}, \frac{2}{m}) = 1$ for $m \in \mathbb{N}$, but $f(\frac{1}{m}, 0) = 0$ for $m \in \mathbb{N}$. This shows that f is *not* jointly continuous at the origin.

Exercise 5.5. Show that the function

$$f(x,y) = \begin{cases} 1 & \text{for } x^2 < y < 3x^2 \\ 0 & \text{otherwise} \end{cases} \tag{61}$$

is *not* jointly continuous at the origin, but *is* continuous along any line, that is, the function $g(t) = f(\alpha + at, \beta + bt)$ is continuous in $\mathbb{R}$ for any constants $\alpha, \beta, a, b \in \mathbb{R}$.

Example 5.6. (a) Let $c \in \mathbb{R}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be the function given by $f(x) = c$ for $x \in \mathbb{R}^n$. Then f is continuous at every point $y \in \mathbb{R}$, since for any sequence $\{x^{(i)}\} \subset \mathbb{R}^n$ converging to y, we have $f(x^{(i)}) = c \to c = f(y)$ as $i \to \infty$.

(b) Let $k \in \{1, \ldots, n\}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be the function given by $f(x) = x_k$ for $x \in \mathbb{R}^n$. Then f is continuous at every point $y \in \mathbb{R}^n$, because given any sequence $\{x^{(i)}\} \subset \mathbb{R}^n$ converging to y, we have $f(x^{(i)}) = x_k^{(i)} \to y_k = f(y)$ as $i \to \infty$.

Following the pattern of the single variable theory, our next step is to combine known continuous functions to create new continuous functions.

Definition 5.7. Given two functions $f, g : K \to \mathbb{R}$, with $K \subset \mathbb{R}^n$, we define their *sum*, *difference*, *product*, and *quotient* by

$$(f \pm g)(x) = f(x) \pm g(x), \qquad (fg)(x) = f(x)g(x), \qquad \text{and} \qquad \big(\frac{f}{g}\big)(x) = \frac{f(x)}{g(x)}, \tag{62}$$

for $x \in K$, where for the quotient definition we assume that $g(x) \neq 0$ for all $x \in K$.

Lemma 5.8. *Let $K \subset \mathbb{R}^n$, and let $f, g : K \to \mathbb{R}$ be functions continuous at $x \in K$. Then the sum and difference $f \pm g$, and the product fg are all continuous at x. Moreover, the function $\frac{1}{f}$ is continuous at x, provided that $f(x) \neq 0$.*

Proof. The results are immediate from the definition of continuity. For instance, let us prove that fg is continuous at x. Thus let $\{x^{(i)}\} \subset K$ be an arbitrary sequence converging to x. Then $f(x^{(i)}) \to f(x)$ and $g(x^{(i)}) \to g(x)$ as $i \to \infty$, and hence $f(x^{(i)})g(x^{(i)}) \to f(x)g(x)$ as $i \to \infty$. Therefore fg is continuous at x. □

Exercise 5.9. Complete the proof of the preceding lemma.

Example 5.10. (a) Recall from Example 5.6 that the constant function $f(x) = c$ (where $c \in \mathbb{R}$) and the projection map $f(x) = x_k$ are continuous in $\mathbb{R}^n$. Then by Lemma 5.8, any monomial $f(x) = ax_1^{\alpha_1} \cdots x_n^{\alpha_n}$ with a constant $a \in \mathbb{R}$, and indices $\alpha_1, \ldots, \alpha_n \in \mathbb{N}_0$, is continuous in $\mathbb{R}^n$, since we can write $ax^n = a \cdot x_1 \cdots x_1 x_2 \cdots x_n$. An n-variable *polynomial* is a function $p : \mathbb{R}^n \to \mathbb{R}$ of the form

$$p(x) = \sum_{\alpha_1,\ldots,\alpha_n} a_{\alpha_1\ldots\alpha_n} x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \tag{63}$$

where only finitely many of the coefficients $a_{\alpha_1\ldots\alpha_n} \in \mathbb{R}$ are nonzero. Applying Lemma 5.8 again, we conclude that all n-variable polynomials are continuous in $\mathbb{R}^n$.

(b) Let p and q be polynomials, and let $Z = \{x \in \mathbb{R}^n : q(x) = 0\}$ be the set of zeroes of q. Then by Lemma 5.8, the function $r : \mathbb{R}^n \setminus Z \to \mathbb{R}$ given by $r(x) = \frac{p(x)}{q(x)}$ is continuous in $\mathbb{R}^n \setminus Z$. The functions of this form are called *rational functions*. For instance, $f(x, y) = \frac{x^2+1}{(x-1)^2+y^2}$ is continuous at each $(x, y) \in \mathbb{R}^2 \setminus \{(1, 0)\}$.

Exercise 5.11. Let $K \subset \mathbb{R}^n$, and let $g : K \to \mathbb{R}^m$ be function whose components are all continuous at $x \in K$. Suppose that $U \subset \mathbb{R}^m$ satisfies $g(K) \subset U$, the latter meaning that $y \in K$ implies $g(y) \in U$. Let $F : U \to \mathbb{R}$ be a function continuous at $g(x)$. Then prove that the composition $F \circ g : K \to \mathbb{R}$, defined by $(F \circ g)(y) = F(g(y))$, is continuous at x.

Exercise 5.12. (a) Show that $f(x, y) = \cos(2x + y) - \sin x$ is continuous in $\mathbb{R}^2$.

(b) Let $K \subset \mathbb{R}^n$ and let $f : K \to \mathbb{R}$ be continuous at $y \in K$. Show that the function $|f|$ defined by

$$|f|(z) = |f(z)| \qquad \text{for} \quad x \in K, \tag{64}$$

is continuous at y.

(c) Show that functions of the form $\frac{r_1(x)+r_2(|x|_\infty)}{r_3(x)+r_4(|x|_\infty)}$ are continuous in an appropriate subset of $\mathbb{R}^n$, where r_1, r_2, r_3, and r_4 are all rational functions.

The following is the ε-δ criterion we mentioned earlier.

Lemma 5.13. *Let $K \subset \mathbb{R}^n$ be a set, and let $f : K \to \mathbb{R}$ be a function. Then f is continuous at $y \in K$ if and only if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $x \in K$ and $0 < |x - y|_\infty < \delta$ imply $|f(x) - f(y)| < \varepsilon$.*

Proof. We first prove the "if" part of the lemma. Assume that the latter condition holds, and let $\{x^{(i)}\} \subset K \setminus \{y\}$ be a sequence converging to y. We want to show that $f(x^{(i)}) \to f(y)$ as $i \to \infty$. Let $\varepsilon > 0$ be arbitrary. Then by assumption, there exists $\delta > 0$ such that $0 < |x - y|_\infty < \delta$ and $x \in K$ imply $|f(x) - f(y)| < \varepsilon$. Since $x^{(i)} \to y$ as $i \to \infty$, there is N such that $|x^{(i)} - y|_\infty < \delta$ whenever $i > N$. Hence we have $|f(x^{(i)}) - f(y)| < \varepsilon$ whenever $n > N$. As $\varepsilon > 0$ is arbitrary, we conclude that $f(x^{(i)}) \to f(y)$ as $i \to \infty$.

To prove the other direction, assume the opposite of the conclusion, i.e., that there is some $\varepsilon > 0$, such that for any $\delta > 0$, there exists some $x \in K$ with $0 < |x - y|_\infty < \delta$ and $|f(x) - f(y)| \geq \varepsilon$. In particular, taking $\delta = \frac{1}{i}$, we infer the existence of a sequence $\{x^{(i)}\} \subset K$ satisfying $0 < |x^{(i)} - y|_\infty < \frac{1}{n}$, with $|f(x^{(i)}) - f(y)| \geq \varepsilon$ for all i. Thus we have a sequence $\{x^{(i)}\} \subset K \setminus \{y\}$ converging to y, with $f(x^{(i)}) \not\to f(y)$ as $i \to \infty$. □

We now consider *vector functions* of several variables. All that has been said extends to this situation in a straightforward, "componentwise" way.

Definition 5.14. Let $K \subset \mathbb{R}^n$ be a set, and let $f : K \to \mathbb{R}^m$ be a function. We say that f is *continuous at* $y \in K$, if each component of f is continuous at y. If $f : K \to \mathbb{R}^m$ is continuous at each point of K, we say that f is *continuous in* K. The set of all continuous functions in K is denoted by $\mathscr{C}(K, \mathbb{R}^m)$, or simply by $\mathscr{C}(K)$ if the target space $\mathbb{R}^m$ is clear from the context.

Example 5.15. (a) The function $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (x \cos y, y \sin x)$ is obviously continuous in $\mathbb{R}^2$. Hence we have $f \in \mathscr{C}(\mathbb{R}^2, \mathbb{R}^2)$.

(b) Let $f(x) = (\theta(x), x^2)$, where θ is the Heaviside step function, and consider the restriction $g = f|_{(0,2]}$. Then notice that $g(x) = (1, x^2)$ for $(0, 2]$, and $g \in \mathscr{C}((0, 1], \mathbb{R}^2)$.

(c) Let $A \in \mathbb{R}^{m\times n}$ be a matrix and define $f : \mathbb{R}^n \to \mathbb{R}^m$ by $f(x) = Ax$. Then $f \in \mathscr{C}(\mathbb{R}^n, \mathbb{R}^m)$, because the k-th component of f is

$$f_k(x) = a_{k1}x_1 + \ldots + a_{kn}x_n, \tag{65}$$

where a_{ki} are the entries of A.

Exercise 5.16. Show that if $f, g \in \mathscr{C}(K, \mathbb{R}^m)$ and $A \in \mathscr{C}(K, \mathbb{R}^{q\times m})$, then $f \pm g \in \mathscr{C}(K, \mathbb{R}^m)$ and $Af \in \mathscr{C}(K, \mathbb{R}^q)$.

Exercise 5.17. Prove the analogue of Lemma 5.13 for vector valued functions $f : K \to \mathbb{R}^m$.

Before closing this section, we extend the notion of the limit of a function to several variable functions, and leave the task of establishing a continuity criterion based on limits of functions as an exercise to the reader.

Definition 5.18. Let $K \subset \mathbb{R}^n$ be a set, and let $f : K \to \mathbb{R}^m$. We say that $f(x)$ *converges to* $\alpha \in \mathbb{R}^m$ *as* $x \to y \in \mathbb{R}^n$, and write

$$f(x) \to \alpha \qquad \text{as} \quad x \to y, \qquad \text{or} \qquad \lim_{x\to y} f(x) = \alpha, \tag{66}$$

if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - \alpha|_\infty < \varepsilon$ whenever $0 < |x - y|_\infty < \delta$ and $x \in K$. One can write $\lim\limits_{x\in K,\, x\to y} f(x)$, $K \ni x \to y$, etc., to explicitly indicate the domain K.

Exercise 5.19. Let $K \subset \mathbb{R}^n$ and let $f : K \to \mathbb{R}^m$. Prove the following.

(a) $f(x) \to \alpha \in \mathbb{R}^m$ as $x \to y \in \mathbb{R}^n$ if and only if $f(x^{(i)}) \to \alpha$ as $i \to \infty$ for every sequence $\{x^{(i)}\} \setminus \{y\} \subset K$ converging to y.

(b) f is continuous at $y \in K$ if and only if $f(x) \to f(y)$ as $x \to y$.

In particular, from the preceding exercise, we see that continuity of f at y is equivalent to $f(x) = f(y) + e(x)$, with $|e(x)|_\infty \to 0$ as $x \to y$, that is, $f(x)$ is approximated by the constant vector $f(y)$ with the error of $o(1)$. To make it precise, we need to extend the little 'o' notation to functions of several variables.

Definition 5.20 (Little 'o' notation)**.** Let $K \subset \mathbb{R}^n$ be a set, let $y \in \mathbb{R}^n$, and let $f : K \to \mathbb{R}^n$ and $g : K \to \mathbb{R}$. Then we write

$$f(x) = o(g(x)) \qquad \text{as} \quad x \to y, \tag{67}$$

to mean that

$$\frac{|f(x)|_\infty}{|g(x)|} \to 0 \qquad \text{as} \quad x \to y. \tag{68}$$

Furthermore, for $h : K \to \mathbb{R}^n$, the notation

$$f(x) = h(x) + o(g(x)) \qquad \text{as} \quad x \to y, \tag{69}$$

is understood to be

$$f(x) - h(x) = o(g(x)) \qquad \text{as} \quad x \to y. \tag{70}$$

Remark 5.21. Continuity of f at y is equivalent to $f(x) = f(y) + o(1)$ as $x \to y$.

6. Differentiability

Separate continuity of a function $f : \mathbb{R}^2 \to \mathbb{R}$ at $y \in \mathbb{R}^2$ is defined in terms of continuity of the single variable functions $g_1(t) = f(y_1 + t, y_2)$ and $g_2(t) = f(y_1, y_2 + t)$, at $t = 0$. Observe that these two functions can also be written as $g_k(t) = f(y + e_k t)$, $k = 1, 2$, where $e_1 = (1, 0)$ and $e_2 = (0, 1)$. In other words, g_k is simply f restricted to the line $\gamma_k(t) = y + e_k t$, $t \in \mathbb{R}$. Apart from continuity, we can talk about differentiability of g_1 and g_2, leading to the notion of partial derivatives. More generally, given an arbitrary vector $V \in \mathbb{R}^2$, the restriction $g(t) = f(y + Vt)$ to the line $\gamma(t) = y + Vt$ can be considered. This leads us to the notion of directional derivatives. Similarly to the situation with continuity, partial and directional derivatives turn out to be *not* the correct generalization of the derivative to higher dimensions, but will be a very useful auxiliary tool to get a handle on the ultimate generalization.

Definition 6.1. Let $K \subset \mathbb{R}^n$. We define the *directional derivative of* $f : K \to \mathbb{R}^m$ *at* $x \in K$ *along* $V \in \mathbb{R}^n$, to be $D_V f(x) = g'(0)$ if the latter exists, where

$$g(t) = f(x + Vt), \tag{71}$$

is a function of $t \in \mathbb{R}$ with $x + Vt \in K$. The *k-th partial derivative of f at x* is simply

$$\partial_k f(x) = \frac{\partial f}{\partial x_k}(x) = D_{e_k} f(x), \tag{72}$$

provided that it exists, where $e_k \in \mathbb{R}^n$ is defined by $(e_k)_i = 0$ for $i \neq k$ and $(e_k)_k = 1$. The matrix consisting of the partial derivatives

$$J(x) = \begin{pmatrix} \partial_1 f_1(x) & \partial_2 f_1(x) & \dots & \partial_n f_1(x) \\ \partial_1 f_2(x) & \partial_2 f_2(x) & \dots & \partial_n f_2(x) \\ \dots & \dots & \dots & \dots \\ \partial_1 f_m(x) & \partial_2 f_m(x) & \dots & \partial_n f_m(x) \end{pmatrix} \tag{73}$$

is called the *Jacobian matrix* of f at x.

Example 6.2. (a) Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} 1 & \text{for } x^2 < y < 3x^2 \\ 0 & \text{otherwise.} \end{cases} \tag{74}$$

Given any $(a, b) \in \mathbb{R}^2$, we have $f(at, bt) = 0$ for all $t > 0$ sufficiently small. Hence the directional derivative $D_V f(0, 0)$ exists and is equal to 0 for all $V \in \mathbb{R}^2$. In particular, the partial derivatives are

$$\frac{\partial f}{\partial x}(0, 0) = D_{(1,0)} f(0, 0) = 0, \qquad \frac{\partial f}{\partial y}(0, 0) = D_{(0,1)} f(0, 0) = 0, \tag{75}$$

and thus the Jacobian matrix of f at the origin is given by $J = (0\ 0) \in \mathbb{R}^{1\times 2}$. However, f is not even continuous at the origin.

(b) Similarly, let

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{for } |x| + |y| > 0 \\ 0 & \text{for } x = y = 0. \end{cases} \tag{76}$$

For $(a, b) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ and $t \neq 0$, we have

$$g(t) = f(at, bt) = \frac{a^2 bt}{a^4 t^2 + b^2} = f(0, 0) + \frac{a^2 b}{a^4 t^2 + b^2} \cdot t, \tag{77}$$

which implies that $g'(0)$ exists, with $g'(0) = a^2/b$ for $b \neq 0$ and $g'(0) = 0$ for $b = 0$. Hence, the directional derivative $D_{(a,b)} f(0, 0)$ exists, with its value equal to a^2/b for $b \neq 0$ and 0 for

$b = 0$. Note that the value a^2/b diverges as $(a, b) \to (1, 0)$, even though $D_{(1,0)}f(0, 0) = 0$, meaning that the dependence of $D_{(a,b)}f(0, 0)$ on (a, b) is *not* continuous. The Jacobian matrix of f at the origin is given by $J = (0\ 0) \in \mathbb{R}^{1\times 2}$.

(c) It is easy to see that the function $f(x, y) = \sqrt{|xy|}$ is differentiable at $(0, 0)$ along V if and only if $V = (a, 0)$ or $V = (0, a)$ for some $a \in \mathbb{R}$.

Remark 6.3. There is no obvious *a priori* structure on how $D_V f$ depends on V, except to say that $D_V f(x)$ is homogeneous in V, that is, $D_{\alpha V} f(x) = \alpha D_V f(x)$ for $\alpha \in \mathbb{R}$.

Exercise 6.4. Let $Q = (a_1, b_1) \times \ldots \times (a_n, b_n) \subset \mathbb{R}^n$ be an n-dimensional rectangular domain, and let $f : Q \to \mathbb{R}$. Pick $y \in Q$ and $V \in \mathbb{R}^n$. Prove the following.

(a) There exists $\varepsilon > 0$ such that the function $g(t) = f(y + Vt)$ is defined for all $t \in (-\varepsilon, \varepsilon)$.

(b) If $D_V f(x)$ exists at each $x \in Q$, then $g'(t) = D_V f(y + Vt)$ as long as t is such that $g(t)$ is well defined, where $g(t)$ is as in the preceding item.

Example 6.5. Let $f : \mathbb{R}^n \to \mathbb{R}^{k\times\ell}$ be a *matrix-valued* function. As an example, $f(x) = xx^T$, that is, $f_{ab}(x) = x_a x_b$ in components, would be $f : \mathbb{R}^n \to \mathbb{R}^{n\times n}$, and its partial derivatives are

$$\frac{\partial f_{ab}}{\partial x_c}(x) = \frac{\partial(x_a x_b)}{\partial x_c} = \delta_{ac}x_b + \delta_{bc}x_a, \tag{78}$$

where δ_{ik} is the Kronecker delta. The Jacobian matrix of such a function would be a matrix of dimension $m\times n$, with $m = k\ell$, because the space $\mathbb{R}^{k\times\ell}$ of $k \times \ell$ matrices is naturally identified with the vector space $\mathbb{R}^m$ with $m = k\ell$. Thus a typical element of the Jacobian matrix would be $J_{i,j} \in \mathbb{R}$, with $i \in \{1, \ldots, k\ell\}$ and $j \in \{1, \ldots, n\}$. However, it might be more convenient to index the elements of the Jacobian matrix as $J_{a,b;j} \in \mathbb{R}$, with $a \in \{1, \ldots, k\}$, $b \in \{1, \ldots, \ell\}$ and $j \in \{1, \ldots, n\}$, since it is a better reflection of the structure of the matrix space $\mathbb{R}^{k\times\ell}$. If we follow this convention, then the rows of the Jacobian matrix would be indexed by two indices, and hence in a certain sense, there would be a "rectangular" set of indices in the row direction, instead of an "interval" of indices.

Exercise 6.6. Following the spirit of the preceding example, discuss what would be the Jacobian matrix of a function of the form $f : \mathbb{R}^{m\times n} \to \mathbb{R}^{k\times\ell}$. Compute the Jacobian matrix of the function $f : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times n}$ given by $f(A) = A^T A$.

Loosely speaking, the way we defined partial derivatives resembles that of separate continuity. Now we want to introduce a notion of derivative that mirrors joint continuity. To motivate it, recall that $f : \mathbb{R} \to \mathbb{R}^m$ is differentiable at y if and only if

$$f(x) = f(y) + \lambda(x - y) + o(x - y) \qquad \text{as} \quad x \to y, \tag{79}$$

for some (fixed) vector $\lambda \in \mathbb{R}^m$. In a certain sense, differentiable functions are well approximated locally by linear functions. The following natural extension of this criterion to functions of several variables was first studied by Karl Weierstrass (1861), Otto Stolz (1893), William H. Young (1910), and Maurice Fréchet (1911).

Definition 6.7. Let $K \subset \mathbb{R}^n$. A function $f : K \to \mathbb{R}^m$ is called *differentiable* at $y \in K$ if

$$f(x) = f(y) + \Lambda(x - y) + o(|x - y|_\infty), \qquad \text{as} \quad x \to y, \tag{80}$$

for some matrix $\Lambda \in \mathbb{R}^{m\times n}$. We call $Df(y) = \Lambda$ if it exists, the *derivative* of f at y.

Remark 6.8. In (80), for a matrix $\Lambda \in \mathbb{R}^{m\times n}$ and a vector $h \in \mathbb{R}^n$, the product $\Lambda h \in \mathbb{R}^m$ is given by

$$(\Lambda h)_i = \Lambda_{i1}h_1 + \ldots + \Lambda_{in}h_n, \qquad i = 1, \ldots, m, \tag{81}$$

where $\Lambda_{ik} \in \mathbb{R}$ is the element in the i-th row and k-th column of Λ. Thus we are implicitly identifying the elements of $\mathbb{R}^n$ and those of $\mathbb{R}^m$ with *column vectors*. However, in general,

$Df(y) = \Lambda$ should be considered as nothing more than a *linear map* $\Lambda : \mathbb{R}^n \to \mathbb{R}^m$. The action Λh can still be defined by (81), but the input h and the output $w = \Lambda h$ are simply vectors in $\mathbb{R}^n$ and in $\mathbb{R}^m$, respectively, rather than numbers arranged in a row or column. This point of view is natural, for instance, when one wants to differentiate functions of the form $f : \mathbb{R}^n \to \mathbb{R}^{k\times\ell}$ or $f : \mathbb{R}^{m\times n} \to \mathbb{R}^{k\times\ell}$.

Note in particular that if f is differentiable at $y \in K$ then f is continuous at y. In contrast, recall from Example 6.2 that directional differentiability does not imply continuity.

Remark 6.9. Suppose that $f : K \to \mathbb{R}^m$ is differentiable at $y \in K$. Then for $V \in \mathbb{R}^n$ fixed and $t \in \mathbb{R}$ with $t \to 0$, we have

$$g(t) := f(y + Vt) = f(y) + t\Lambda V + o(t), \tag{82}$$

where $\Lambda = Df(y)$, and we have taken into account $o(|Vt|_\infty) = o(t)$. This leads to

$$\frac{g(t) - g(0)}{t} = \Lambda V + o(1) \to \Lambda V \qquad \text{as} \quad t \to 0, \tag{83}$$

i.e., the directional derivative $D_V f(y)$ exists, with $D_V f(y) = \Lambda V = Df(y)V$. Thus differentiability of f implies not only directional differentiability, but also a linear (and hence continuous) dependence of the derivative $D_V f(y)$ on the direction V. In particular, by taking $V = e_k$, we see that the derivative $Df(y)$ is in fact equal to the Jacobian matrix $J(y)$ of f.

In view of the preceding remark, if we somehow know that f is differentiable, we can compute the derivative as the Jacobian matrix. However, how do we ascertain differentiability of f in the first place? The following result gives a practical way to handle this situation.

Theorem 6.10. *Let $Q = (a_1, b_1) \times \ldots \times (a_n, b_n)$ be an n-dimensional rectangular domain, and let $f : Q \to \mathbb{R}^m$. Suppose that all partial derivatives of f exist at each $x \in Q$, and that the partial derivatives are continuous in Q. Then f is differentiable in Q.*

Proof. We will prove only the case $n = 2$ and $m = 1$, as this is the simplest case that can illustrate all the essential ideas. For $x, y \in Q$, we have

$$f(x) - f(y) = f(x_1, x_2) - f(y_1, x_2) + f(y_1, x_2) - f(y_1, y_2). \tag{84}$$

By applying the mean value theorem to $g_1(t) = f(t, x_2)$ and to $g_2(t) = f(y_1, t)$, we infer the existence of $\xi \in [y_1, x_1] \cup [x_1, y_1]$ and $\eta \in [y_2, x_2] \cup [x_2, y_2]$, satisfying

$$f(x) - f(y) = \partial_1 f(\xi, x_2)(x_1 - y_1) + \partial_2 f(y_1, \eta)(x_2 - y_2). \tag{85}$$

Since both $\partial_1 f : Q \to \mathbb{R}$ and $\partial_2 f : Q \to \mathbb{R}$ are continuous, we have $\partial_1 f(\xi, x_2) \to \partial_1 f(y)$ and $\partial_2 f(y_1, \eta) \to \partial_2 f(y)$ as $x \to y$. In other words, we have

$$f(x) - f(y) = \partial_1 f(y)(x_1 - y_1) + \partial_2 f(y)(x_2 - y_2) + h_1(x)(x_1 - y_1) + h_2(x)(x_2 - y_2), \tag{86}$$

with both $h_1(x) \to 0$ and $h_2(x) \to 0$ as $x \to y$. As

$$|h_1(x)(x_1 - y_1) + h_2(x)(x_2 - y_2)| \leq (|h_1(x)| + |h_2(x)|)\, |x - y|_\infty, \tag{87}$$

we conclude that

$$\begin{aligned} f(x) - f(y) &= \partial_1 f(y)(x_1 - y_1) + \partial_2 f(y)(x_2 - y_2) + o(|x - y|_\infty) \\ &= \Lambda(x - y) + o(|x - y|_\infty) \qquad \text{as} \quad x \to y, \end{aligned} \tag{88}$$

where $\Lambda = (\partial_1 f(y), \partial_2 f(y)) \in \mathbb{R}^{1\times 2}$, establishing that f is differentiable at y. $\square$

Exercise 6.11. Write out a detailed proof of the preceding lemma in the general case.

Example 6.12. (a) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x, y) = (x \cos y, y \sin x)$. Its Jacobian matrix can be computed as

$$J(x, y) = \begin{pmatrix} \partial_x f & \partial_y f \end{pmatrix} = \begin{pmatrix} \cos y & -x \sin y \\ y \cos x & \sin x \end{pmatrix}. \tag{89}$$

Since $J : \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$ is continuous in $\mathbb{R}^2$, we conclude that f is differentiable in $\mathbb{R}^2$ with $Df(x) = J(x)$.

(b) Let $A \in \mathbb{R}^{m\times n}$ be a matrix and define $f : \mathbb{R}^n \to \mathbb{R}^m$ by $f(x) = Ax$. In components, it is

$$f_j(x) = a_{j1}x_1 + \ldots + a_{jn}x_n, \tag{90}$$

where a_{jk} are the entries of A, and thus $\partial_k f_j(x) = a_{jk}$. This means that the Jacobian matrix of f is the matrix $J(x) = A$, independent of x. Since constant functions are continuous, f is differentiable in $\mathbb{R}^n$ with $Df(x) = A$.

Exercise 6.13. Prove the following.

(a) $|x + y|_\infty \leq |x|_\infty + |y|_\infty$ for $x, y \in \mathbb{R}^n$.

(b) $|Ax|_\infty \leq C|x|_\infty$ for $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, where the constant C may depend only on A.

Lemma 6.14 (Chain rule)**.** *Let $K \subset \mathbb{R}^n$ be a set, and let $f : K \to \mathbb{R}^m$ be a function, differentiable at $y \in K$. Suppose that $U \subset \mathbb{R}^m$, such that $f(K) \subset U$, that is, $f(x) \in U$ for all $x \in K$. Assume that $g : U \to \mathbb{R}^k$ is differentiable at $f(y)$. Then the composition $g \circ f : K \to \mathbb{R}^k$, defined by $(g \circ f)(x) = g(f(x))$ for $x \in K$, is differentiable at y, with*

$$D(g \circ f)(y) = Dg(f(y))Df(y). \tag{91}$$

Proof. By definition, we have

$$f(x) = f(y) + A(x - y) + h(x) \qquad \text{with} \quad h(x) = o(|x - y|_\infty) \quad \text{as} \quad x \to y, \tag{92}$$

where $A = Df(y) \in \mathbb{R}^{m\times n}$, and

$$g(u) = g(f(y)) + B(u - f(y)) + e(u) \qquad \text{with} \quad e(u) = o(|u - f(y)|_\infty) \quad \text{as} \quad u \to f(y), \tag{93}$$

where $B = Dg(f(y)) \in \mathbb{R}^{k\times m}$. Putting $u = f(x)$ in the latter formula, we get

$$\begin{aligned} g(f(x)) &= g(f(y)) + B(f(x) - f(y)) + e(f(x)) \\ &= g(f(y)) + BA(x - y) + Bh(x) + e(f(x)). \end{aligned} \tag{94}$$

The proof is complete upon showing that $Bh(x) + e(f(x)) = o(|x - y|_\infty)$ as $x \to y$. Indeed, there is some constant C such that $|Bh(x)|_\infty \leq C|h(x)|_\infty$, which implies that

$$\frac{|Bh(x)|_\infty}{|x - y|_\infty} \leq C\frac{|h(x)|_\infty}{|x - y|_\infty} \to 0 \qquad \text{as} \quad x \to y. \tag{95}$$

For the other term, we have

$$\frac{|e(f(x))|_\infty}{|x - y|_\infty} = \frac{|e(f(x))|_\infty}{|f(x) - f(y)|_\infty}\frac{|f(x) - f(y)|_\infty}{|x - y|_\infty} \to 0 \qquad \text{as} \quad x \to y, \tag{96}$$

because

$$\frac{|f(x) - f(y)|_\infty}{|x - y|_\infty} \leq \frac{|A(x - y)|_\infty + |h(x)|_\infty}{|x - y|_\infty} \leq C' + \frac{|h(x)|_\infty}{|x - y|_\infty}. \tag{97}$$

This completes the proof. ☐

Exercise 6.15. In the context of Definition 6.7, show that f is differentiable at $y \in K$ if and only if there exists a function $G : K \to \mathbb{R}^{m\times n}$, continuous at y, such that

$$f(x) = f(y) + G(x)(x - y) \qquad \text{for} \quad x \in K. \tag{98}$$

This is of course an extension of Carathéodory's criterion.

7. Second order derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable everywhere in $\mathbb{R}^n$. By the convention that the elements of $\mathbb{R}^n$ are column vectors, it is natural to consider $Df(x)$ as a row vector, that is, $Df(x) \in \mathbb{R}^{1\times n}$ for $x \in \mathbb{R}^n$. Hence, Df can be considered as a function $Df : \mathbb{R}^n \to \mathbb{R}^{1\times n}$, and we can talk about its differentiability. Then the derivative $D^2 f(y) = DDf(y)$, if exists, should be a linear map $\Lambda : \mathbb{R}^n \to \mathbb{R}^{1\times n}$, satisfying

$$Df(x) = Df(y) + \Lambda(x - y) + o(|x - y|_\infty) \qquad \text{as} \quad x \to y. \tag{99}$$

Now, any such map can be written as

$$\Lambda(h) = h^T A, \qquad h \in \mathbb{R}^n, \tag{100}$$

for some matrix $A \in \mathbb{R}^{n\times n}$. In view of this, we are going to identify $D^2 f(x)$ with the matrix A, and write

$$Df(x) = Df(y) + (x - y)^T D^2 f(y) + o(|x - y|_\infty) \qquad \text{as} \quad x \to y. \tag{101}$$

Taking the transpose of this equation, we get the equivalent form

$$Df(x)^T = Df(y)^T + D^2 f(y)^T (x - y) + o(|x - y|_\infty) \qquad \text{as} \quad x \to y, \tag{102}$$

which means that $D^2 f(y)^T$ corresponds to the Jacobian matrix of the function $(Df)^T$ at y.

Definition 7.1. Let $K \subset \mathbb{R}^n$, and let $f : K \to \mathbb{R}$ be differentiable everywhere in K. If $Df : K \to \mathbb{R}^{1\times n}$ is differentiable at $y \in K$ in the sense (101), then we say that f is *twice differentiable at* y, and call $D^2 f \in \mathbb{R}^{n\times n}$ the *second order derivative of* f *at* y.

Remark 7.2. In light of (102), f is twice differentiable at y if and only if there is a matrix $H \in \mathbb{R}^{n\times n}$ such that

$$B(x) = B(y) + H^T (x - y) + o(|x - y|_\infty) \qquad \text{as} \quad x \to y, \tag{103}$$

where $B(x) = Df(x)^T$ is simply the transpose of the Jacobian of f. Note that the Jacobian of f is the row vector consisting of the partial derivatives of f, and so its transpose $B(x)$ is a column vector. Hence H^T is given by the Jacobian matrix of the function $B(x)$, that is,

$$H^T = \begin{pmatrix} \partial_1\partial_1 f(y) & \partial_2\partial_1 f(y) & \dots & \partial_n\partial_1 f(y) \\ \partial_1\partial_2 f(y) & \partial_2\partial_2 f(y) & \dots & \partial_n\partial_2 f(y) \\ \dots & \dots & \dots & \dots \\ \partial_1\partial_n f(y) & \partial_n\partial_1 f(y) & \dots & \partial_n\partial_n f(y) \end{pmatrix}, \tag{104}$$

or

$$H = \begin{pmatrix} \partial_1\partial_1 f(y) & \partial_1\partial_2 f(y) & \dots & \partial_1\partial_n f(y) \\ \partial_2\partial_1 f(y) & \partial_2\partial_2 f(y) & \dots & \partial_2\partial_n f(y) \\ \dots & \dots & \dots & \dots \\ \partial_n\partial_1 f(y) & \partial_n\partial_2 f(y) & \dots & \partial_n\partial_n f(y) \end{pmatrix}. \tag{105}$$

This matrix is called the *Hessian of* f *at* y. Note that if $D^2 f(y)$ exists, then $D^2 f(y) = H$, but without additional assumptions, the existence of H does not imply the existence of $D^2 f(y)$.

Example 7.3. (a) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = ax^2 + 2bxy + cy^2$, where $a, b, c \in \mathbb{R}$ are constants. We can compute the Jacobian matrix of f at (x, y) as

$$J_f(x, y) = \begin{pmatrix} 2ax + 2by & 2bx + 2cy \end{pmatrix} \in \mathbb{R}^{1\times 2}. \tag{106}$$

Since this depends on $(x, y) \in \mathbb{R}^2$ continuously, we conclude that f is differentiable everywhere in $\mathbb{R}^2$, with $Df(x, y) = J_f(x, y)$. Now, the Jacobian matrix of $(Df)^T$ is

$$J(x, y) = \begin{pmatrix} 2a & 2b \\ 2b & 2c \end{pmatrix}, \tag{107}$$

and since it is continuous in $\mathbb{R}^2$, we conclude that f is twice differentiable in $\mathbb{R}^2$ with $D^2 f(x,y) = J^T = J$.

(b) More generally, let $A \in \mathbb{R}^{n\times n}$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by $f(x) = x^T Ax$, i.e.,

$$f(x) = \sum_{i,k=1}^{n} a_{ik} x_i x_k, \tag{108}$$

where $a_{ik} \in \mathbb{R}$ are the elements of A. Note that for $i \neq k$, the combination $x_i x_k$ appears in the sum twice, so that we can in fact write

$$f(x) = \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{i=1}^{n} \sum_{k=1}^{i-1} (a_{ik} + a_{ki}) x_i x_k. \tag{109}$$

In any case, we have

$$\partial_j f(x) = \sum_{i,k=1}^{n} a_{ik} (\delta_{kj} x_i + \delta_{ij} x_k) = \sum_{i=1}^{n} a_{ij} x_i + \sum_{k=1}^{n} a_{jk} x_k, \tag{110}$$

and hence

$$Df(x) = x^T A + x^T A^T = x^T (A + A^T). \tag{111}$$

For the transpose, we would have $Df(x)^T = (A + A^T)x$, and the Jacobian of $(Df)^T$ is simply $J = A + A^T$. The Hessian of f is then the transpose of J, which is $H = A + A^T$. We conclude that f is twice differentiable in $\mathbb{R}^n$ with $D^2 f(x) = A + A^T$. Note that the Hessian is always symmetric no matter what A is, but this would have been clear from (109), which shows that replacing A by $\frac{1}{2}(A + A^T)$ does not change f.

The following is a practical criterion similar to Theorem 6.10 on twice differentiability.

Remark 7.4. Let $Q = (a_1, b_1) \times \ldots \times (a_n, b_n)$ be an n-dimensional rectangular domain, and let $f : Q \to \mathbb{R}$ be a function. Assume that the Hessian $H(x)$ exists at every $x \in Q$, and $H(x)$ depends on x continuously. The existence of the Hessian in particular guarantees the existence of all first order partial derivatives of f in Q. Let $B(x) \in \mathbb{R}^n$ be the (column) vector consisting of all first order partial derivatives of f at x. Then $H(x)^T$ is the Jacobian of the mapping $B : Q \to \mathbb{R}^n$ at x, and by Theorem 6.10, continuity of H implies that B is differentiable in Q, with $DB = H^T$. In particular, B is continuous in Q. Now, since $B(x)^T$ is the Jacobian of f at x, this implies that f is differentiable in Q, with $Df = B^T$. As we have already established that B is differentiable in Q, we conclude that Df is differentiable in Q, and that $D^2 f = H$.

Differentiable functions are well approximated locally by linear functions. Intuitively speaking, twice differentiable functions should be close to quadratic functions. To make it precise, we first recall the corresponding one dimensional result. For completeness, we include a proof.

Theorem 7.5. *a) If $f : (a,b) \to \mathbb{R}$ is twice differentiable at $c \in (a,b)$, then there is a function $\psi : (a,b) \to \mathbb{R}$ that is continuous at c with $\psi(c) = 0$, such that*

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2}(x-c)^2 + \psi(x)(x-c)^2, \qquad x \in (a,b). \tag{112}$$

b) If $f : (a,b) \to \mathbb{R}$ is twice differentiable in (a,b), and $y, c \in (a,b)$, then there exists $\xi \in (y,c) \cup (c,y)$, such that

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(\xi)}{2}(x-c)^2. \tag{113}$$

Proof. a) Assume that f is twice differentiable at c. By definition, this means that f is differentiable in $(c-\varepsilon, c+\varepsilon)$ for some $\varepsilon > 0$, and that there is a function $g : (a,b) \to \mathbb{R}$ that is continuous at c with $g(c) = f''(c)$, such that

$$f'(x) = f'(c) + g(x)(x-c), \qquad x \in (c-\varepsilon, c+\varepsilon). \tag{114}$$

In other words, for any sequence $\{x_n\} \subset (c-\varepsilon, c) \cup (c, c+\varepsilon)$, we have

$$\frac{f'(x_n) - f'(c)}{x_n - c} \to f''(c) \qquad \text{as } n \to \infty. \tag{115}$$

Since $[f(x) - f(c) - f'(c)(x-c)]' = f'(x) - f'(c)$ and $[\frac{1}{2}(x-c)^2]' = x - c$, by L'Hôpital's rule, for any sequence $\{x_n\} \subset (c-\varepsilon, c) \cup (c, c+\varepsilon)$, this implies that

$$\frac{f(x_n) - f(c) - f'(c)(x_n - c)}{\frac{1}{2}(x_n - c)^2} \to f''(c) \qquad \text{as } n \to \infty. \tag{116}$$

Note that the function

$$F(x) = \frac{f(x) - f(c) - f'(c)(x-c)}{\frac{1}{2}(x-c)^2}, \tag{117}$$

is well defined in $(a,c) \cup (c,b)$. Then upon setting $F(c) = f''(c)$, by (116) we have F continuous at c. In other words, there is a function $\psi : (a,b) \to \mathbb{R}$ that is continuous at c with $\psi(c) = 0$, such that

$$f(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2}(x-c)^2 + \psi(x)(x-c)^2, \qquad x \in (a,b). \tag{118}$$

Thus in a certain sense, the existence of the second derivative guarantees that the function can be approximated by a quadratic polynomial well.

b) In the standard proof of the mean value theorem, we construct a function of the form $F(x) = f(x) + A(x-c)$ with $F(c) = F(y) = f(c)$. Here we look for a function $F(x) = f(x) + A(x-c) + B(x-c)^2$ with $F(c) = F(y) = f(c)$ and $F'(c) = 0$, and easily find such a function as

$$F(x) = f(x) - f'(c)(x-c) - \big[f(y) - f(c) - f'(c)(y-c)\big]\frac{(x-c)^2}{(y-c)^2}. \tag{119}$$

Then F is twice differentiable in (c,y), with

$$F'(x) = f'(x) - f'(c) - \big[f(y) - f(c) - f'(c)(y-c)\big]\frac{2(x-c)}{(y-c)^2}, \tag{120}$$

and

$$F''(x) = f''(x) - \frac{2[f(y) - f(c) - f'(c)(y-c)]}{(y-c)^2}. \tag{121}$$

Moreover, $F'(c)$ exists and $F' \in \mathscr{C}([c,y))$. Since $F(c) = F(y)$, by Rolle's theorem, there is $\xi \in (c,y)$ such that $F'(\xi) = 0$. Now recalling that $F'(c) = 0$ and $F' \in \mathscr{C}([c,y))$, another application of Rolle's theorem gives the existence of $\eta \in (c,\xi)$ such that $F''(\eta) = 0$. In other words, we have

$$f(y) = f(c) + f'(c)(y-c) + \frac{1}{2}f''(\eta)(y-c)^2, \tag{122}$$

for some $\eta \in (c,y)$. $\square$

Exercise 7.6. Prove the following.

(a) If $f : (a,b) \to \mathbb{R}$ is n times differentiable at $c \in (a,b)$, then there is a function $\psi : (a,b) \to \mathbb{R}$ that is continuous at c with $\psi(c) = 0$, such that

$$f(x) = f(c) + f'(c)(x-c) + \ldots + \frac{f^{(n)}(c)}{n!}(x-c)^n + \psi(x)(x-c)^n, \qquad x \in (a,b). \tag{123}$$

(b) If $f : (a, b) \to \mathbb{R}$ is n times differentiable in (a, b), and if $x, c \in (a, b)$, then there exists $\xi \in (x, c) \cup (c, x)$, such that

$$f(x) = f(c) + f'(c)(x - c) + \ldots + \frac{f^{(n-1)}(c)}{(n-1)!}(x - c)^{n-1} + \frac{f^{(n)}(\xi)}{n!}(x - c)^n. \tag{124}$$

Remark 7.7. Let $f : K \to \mathbb{R}$ with $K \subset \mathbb{R}^n$, and for $V \in \mathbb{R}^n$ fixed, suppose that $D_V f(x)$ exists at each $x \in K$. Then the dependence of $D_V f(x)$ on x can naturally be considered as a function $D_V f : K \to \mathbb{R}$. Hence one can talk about its directional differentiability, that is, about whether $D_W D_V f(x)$ exists for $x \in K$ and $W \in \mathbb{R}^n$. Now suppose that f is twice differentiable at y, i.e.,

$$Df(x) = Df(y) + (x - y)^T D^2 f(y) + o(|x - y|_\infty) \qquad \text{as} \quad x \to y. \tag{125}$$

If we multiply this from the right by $V \in \mathbb{R}^n$, we get

$$D_V f(x) = D_V f(y) + (x - y)^T D^2 f(y) V + o(|x - y|_\infty) \qquad \text{as} \quad x \to y, \tag{126}$$

where we have taken into account that $D_V f(x) = Df(x)V$. Next, we put $x = y + tW$ with $W \in \mathbb{R}^n$ and $t \in \mathbb{R}$ small, and infer

$$D_V f(y + tW) = D_V f(y) + tW^T D^2 f(y) V + o(t) \qquad \text{as} \quad x \to y. \tag{127}$$

This implies that

$$D_W D_V f(y) = W^T D^2 f(y) V. \tag{128}$$

Remark 7.8. Let $Q \subset \mathbb{R}^n$ be an n-dimensional rectangular domain, and let $f : Q \to \mathbb{R}$ be twice differentiable everywhere in Q. Fix some $y \in K$ and $V \in \mathbb{R}^n$, and consider the function $g(t) = f(y + tV)$. We know that

$$\begin{aligned} g'(t) &= D_V f(y + tV) = Df(y + tV)V, \\ g''(t) &= D_V^2 f(y + tV) = V^T D^2 f(y + tV)V, \end{aligned} \tag{129}$$

provided that $D^2 f(y + tV)$ exists. Invoking Theorem 7.5a), we get

$$f(y + tV) = f(y) + tDf(y)V + \frac{t^2}{2} V^T D^2 f(y) V + o(t^2) \qquad \text{as} \quad h \to 0. \tag{130}$$

So not surprisingly, the existence of $D^2 f(y)$ guarantees that $f(y + tV)$ can be well approximated by a quadratic polynomial in t. Supposing that $y + V \in Q$, we can also apply Theorem 7.5b), which yields

$$f(y + V) = f(y) + Df(y)V + \frac{1}{2} V^T D^2 f(y + sV)V, \tag{131}$$

for some $s \in (0, 1)$. This gives a quantitative information on the size of the error of the linear approximation $f(y + V) \approx f(y) + Df(y)V$.

Exercise 7.9. Introduce the notion of third derivative $D^3 f$, and give a quantitative information on the error of the quadratic approximation

$$f(y + V) \approx f(y) + Df(y)V + \frac{1}{2} V^T D^2 f(y) V, \tag{132}$$

in the spirit of (131).

The following is a famous theorem which says that under some assumptions on f, directional differentiations commute with each other. The latter would in particular imply that the Hessian matrix is symmetric. An intuitive explanation of this phenomenon is that twice differentiable functions can be well approximated by quadratic functions, and a quadratic function has a symmetric Hessian, cf. Example 7.3.

Theorem 7.10. *Let $Q \subset \mathbb{R}^n$ be an n-dimensional rectangular domain, and let $f : Q \to \mathbb{R}$. Suppose that for $V, W \in \mathbb{R}^n$, $D_V D_W f$ and $D_W D_V f$ exist and are continuous in Q. Then $D_V D_W f = D_W D_V f$ in Q.*

Proof. Fix $x \in Q$, and $V, W \in \mathbb{R}^n$. For $t, s \in \mathbb{R}$, let

$$\begin{aligned} g(s) &= f(x + tV + sW) - f(x + sW), \\ h(t) &= f(x + tV + sW) - f(x + tV). \end{aligned} \tag{133}$$

Notice that

$$g(s) - g(0) = f(x + tV + sW) - f(x + sW) - f(x + tV) + f(x) = h(t) - h(0). \tag{134}$$

Now taking into account that

$$\begin{aligned} g'(s) &= D_W f(x + tV + sW) - D_W f(x + sW), \\ h'(t) &= D_V f(x + tV + sW) - D_V f(x + tV), \end{aligned} \tag{135}$$

by the mean value theorem, there exist $s_1 \in (s, 0) \cup (0, s)$ and $t_1 \in (t, 0) \cup (0, t)$ such that $g(s) - g(0) = g'(s_1)$ and $h(t) - h(0) = h'(t_1)$. That is, we have

$$D_W f(x + tV + s_1 W) - D_W f(x + s_1 W) = D_V f(x + t_1 V + sW) - D_V f(x + t_1 V) \tag{136}$$

Thinking of the left hand side as a function of t, and the right hand side as a function of s, and applying the mean value theorem once again, we infer

$$D_V D_W f(x + t_2 V + s_1 W) = D_W D_V f(x + t_1 V + s_2 W), \tag{137}$$

for some $s_2 \in (s, 0) \cup (0, s)$ and $t_2 \in (t, 0) \cup (0, t)$. As both $D_V D_W f$ and $D_W D_V f$ are continuous, by sending $s \to 0$ and $t \to 0$, we conclude that $D_V D_W f(x) = D_W D_V f(x)$. $\square$

Exercise 7.11. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} xy & \text{for } -|x| < y < |x|, \\ 0 & \text{otherwise.} \end{cases} \tag{138}$$

Check if the partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ exist at the origin, and if $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ there.

8. Inverse functions: Single variable case

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\alpha \in \mathbb{R}^n$, and consider the equation $f(x) = \alpha$ for the unknown $x \in \mathbb{R}^n$. If this equation has a unique solution for all $\alpha \in U$, where $U \subset \mathbb{R}^n$ is some set, the correspondence $\alpha \mapsto x$ defines the inverse function $f^{-1} : U \to \mathbb{R}^n$ of f on U.

Definition 8.1. Let $K \subset \mathbb{R}^n$ and $U \subset \mathbb{R}^n$, and let $f : K \to U$ be bijective, i.e., for each $\alpha \in U$ there is a unique $x \in K$ such that $f(x) = \alpha$. Then the function $g : U \to K$ defined by $g(f(x)) = x$ for $x \in K$ is called the *inverse function* of f, and denoted by $f^{-1} = g$. More generally, $f : K \to \mathbb{R}^n$ is said to be *invertible in a subset* $A \subset K$ if the restriction $f|_A : A \to V$, defined by $f|_A(x) = f(x)$ for $x \in A$, is invertible, where $V = f(A) \equiv \{f(x) : x \in A\}$.

The *inverse function theorem* is the answer to the invertibility question from a differentiable point of view. To motivate it, suppose that $f : K \to \mathbb{R}^n$ is differentiable at $x^* \in K$, that is, there is $\Lambda \in \mathbb{R}^{n \times n}$ such that

$$f(x) = f(x^*) + \Lambda(x - x^*) + o(|x - x^*|_\infty) \qquad \text{as} \quad x \to x^*. \tag{139}$$

If we define $g : \mathbb{R}^n \to \mathbb{R}^n$ by $g(x) = f(x^*) + \Lambda(x - x^*)$, then $f(x) \approx g(x)$ when x is close to x^*. Moreover, g is invertible if and only if Λ is an invertible matrix. Now, the question is given that g is invertible, can we conclude that f is invertible in some region $A \ni x^*$?

Before studying this question in full generality, it will be very instructive to consider the case $n = 1$. Suppose that $f : (a, b) \to \mathbb{R}$ is differentiable at x^* with $f'(x^*) \neq 0$ for some

$x^* \in (a, b)$. Is it true that f is invertible in $(x^* - r, x^* + r)$ for some $r > 0$? However, the following exercise shows that the answer is negative even if f is differentiable everywhere.

Exercise 8.2. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \frac{1}{2}x + x^2 \sin \frac{1}{x} & \text{for } x \neq 0, \\ 0 & \text{for } x = 0. \end{cases} \tag{140}$$

Show that f is differentiable everywhere, and $f'(0) \neq 0$, but f is not invertible in $(-r, r)$ for any $r > 0$. Is f' continuous at 0?

Thus, we need a stronger assumption, and our updated assumption is that $f : (a, b) \to \mathbb{R}$ is *continuously differentiable* in (a, b), and that $f'(x^*) \neq 0$ for some $x^* \in (a, b)$. Under this assumption, we will be able to prove that f is invertible in $(x^* - r, x^* + r)$ for some $r > 0$. Roughly speaking, we want to show that for every $y \in \mathbb{R}$ near $y^* = f(x^*)$, there is a unique x near x^* such that $f(x) = y$. Since $y \approx y^*$, a good initial guess for x is $x_0 = x^*$. Then we can improve this guess by approximating $f(x)$ by the linear function $g_0(x) = f(x_0) + f'(x^*)(x - x_0)$. In other words, we find x_1 such that $g_0(x_1) = y$, that is,

$$x_1 = x_0 + \frac{y - f(x_0)}{f'(x^*)}. \tag{141}$$

We can repeat this procedure, and define the sequence $\{x_k\}$, where $x_0 = x^*$ and

$$x_{k+1} = x_k + \frac{y - f(x_k)}{f'(x^*)}, \qquad k = 0, 1, \ldots. \tag{142}$$

Note that we can write $x_{k+1} = \phi(x_k)$ with

$$\phi(t) = t + \frac{y - f(t)}{f'(x^*)}. \tag{143}$$

Our hope is that $\{x_k\}$ converges to some x, and that the limit solves the equation $f(x) = y$. To guarantee convergence, we will use *Cauchy's criterion*, which says that if

$$|x_m - x_n| \to 0 \qquad \text{as} \quad \min\{m, n\} \to \infty, \tag{144}$$

then there exists $x \in \mathbb{R}$ such that $x_k \to x$ as $k \to \infty$. As part of this program, we start by studying the function ϕ.

Lemma 8.3. *Given any $\alpha > 0$, there exists $\delta = \delta(\alpha) > 0$ such that*

$$|\phi(s) - \phi(t)| \leq \alpha |s - t| \qquad \textit{for all} \quad s, t \in [x^* - \delta, x^* + \delta]. \tag{145}$$

Moreover, given any $0 < \alpha < 1$ and any $0 < r \leq \delta(\alpha)$, there exists $\varepsilon = \varepsilon(r, \alpha) > 0$ such that $|y - y^*| < \varepsilon$ implies

$$|\phi(t) - x^*| \leq r \qquad \textit{for all} \quad t \in [x^* - r, x^* + r]. \tag{146}$$

Proof. Let $\delta > 0$, whose value will be adjusted later. For $s, t \in [x^* - \delta, x^* + \delta]$, we have

$$\phi(t) - \phi(s) = t - s + \frac{f(s) - f(t)}{f'(x^*)} = \frac{f(s) - f(t) + f'(x^*)(t - s)}{f'(x^*)}. \tag{147}$$

By the mean value theorem, there exists η between s and t, in particular $\eta \in [x^* - \delta, x^* + \delta]$, such that $f(s) - f(t) = f'(\eta)(s - t)$, and hence

$$\phi(t) - \phi(s) = \frac{f'(x^*) - f'(\eta)}{f'(x^*)}(t - s). \tag{148}$$

Now, by continuity of f', we can choose $\delta = \delta(\alpha) > 0$ such that

$$|f'(x^*) - f'(\eta)| \leq \alpha |f'(x^*)| \qquad \text{for any} \quad \eta \in [x^* - \delta, x^* + \delta]. \tag{149}$$

This establishes the first part of the lemma.

As for the second part, we start with

$$\phi(t) - x^* = t - x^* + \frac{y - f(t)}{f'(x^*)} = t - x^* + \frac{f(x^*) - f(t)}{f'(x^*)} + \frac{y - y^*}{f'(x^*)}. \tag{150}$$

Let $0 < \alpha < 1$ and let $0 < r \le \delta(\alpha)$, where $\delta(\alpha) > 0$ is from the first part of this lemma. By the mean value theorem, there is $\eta \in [x^* - r, x^* + r]$ such that $f(x^*) - f(t) = f'(\eta)(x^* - t)$, and hence

$$\phi(t) - x^* = \frac{f'(x^*) - f'(\eta)}{f'(x^*)}(t - x^*) + \frac{y - y^*}{f'(x^*)}. \tag{151}$$

Since $r \le \delta(\alpha)$, the estimate (149) yields

$$|\phi(t) - x^*| \le \alpha|t - x^*| + \frac{|y - y^*|}{|f'(x^*)|}. \tag{152}$$

Then letting $\varepsilon = (1-\alpha)r|f'(x^*)|$, we infer

$$|\phi(t) - x^*| \le \alpha|t - x^*| + \frac{|y - y^*|}{|f'(x^*)|} \le r, \tag{153}$$

whenever $|t - x^*| \le r$ and $|y - y^*| \le \varepsilon$, which completes the proof. □

Now, fix $0 < \alpha < 1$, and $0 < r \le \delta(\alpha)$, and let $\varepsilon = \varepsilon(r, \alpha)$ be as in the lemma. Assume that $y \in [y^* - \varepsilon, y^* + \varepsilon]$. Then for our iteration $x_{k+1} = \phi(x_k)$, $k = 0, 1 \ldots$, with $x_0 = x^*$, we obviously have $x_k \in [x^* - r, x^* + r]$ for all $k = 0, 1, \ldots$. Furthermore, we infer

$$|x_{n+1} - x_n| = |\phi(x_n) - \phi(x_{n-1})| \le \alpha|x_n - x_{n-1}| \le \ldots \le \alpha^n|x_1 - x_0|, \tag{154}$$

and hence

$$\begin{aligned} |x_{n+k} - x_n| &\le |x_{n+k} - x_{n+k-1}| + \ldots + |x_{n+1} - x_n| \le (\alpha^{n+k-1} + \ldots + \alpha^n)|x_1 - x_0| \\ &\le \alpha^n(1 + \alpha + \alpha^2 + \ldots)|x_1 - x_0| = \frac{\alpha^n}{1-\alpha}|x_1 - x_0|, \end{aligned} \tag{155}$$

showing that $\{x_k\}$ is a Cauchy sequence. Thus there exists $x \in \mathbb{R}$ such that $x_k \to x$ as $k \to \infty$.

Let us check if $x \in [x^* - r, x^* + r]$. Suppose that $x \notin [x^* - r, x^* + r]$. Obviously, the point in $[x^* - r, x^* + r]$ that is closest to x would be one of the endpoints, i.e., either $c = x^* - r$ or $d = x^* + r$. This means that $z \in [c, d]$ implies $|x - z| \ge e = \min\{|x - c|, |x - d|\}$, where clearly $e > 0$. In other words, there is a positive distance between the point x and the interval $[c, d]$. Now, since $x_k \to x$, there exists an index n such that $|x - x_n| < e$. But we have $x_n \in [c, d]$, which contradicts with the assertion that $|x - z| \ge e$ for any $z \in [c, d]$. Therefore, our initial assumption $x \notin [c, d]$ is wrong, and we conclude that $x \in [c, d]$.

The next question is if x solves $f(x) = y$. First, observe that if $x = \phi(x)$, i.e., if

$$x = x + \frac{y - f(x)}{f'(x^*)}, \tag{156}$$

then $y = f(x)$. So we will try to show $x = \phi(x)$. For any n, we have

$$\begin{aligned} |x - \phi(x)| &= |x - x_{n+1} + \phi(x_n) - \phi(x)| \le |x - x_{n+1}| + |\phi(x_n) - \phi(x)| \\ &\le |x - x_{n+1}| + \alpha|x_n - x|, \end{aligned} \tag{157}$$

and the right hand side tends to 0 as $n \to \infty$. That is, we have $|x - \phi(x)| \le e$ for any $e > 0$, which means that $|x - \phi(x)| = 0$, and hence $x = \phi(x)$.

Finally, we need to show that x is the only solution of $f(x) = y$ in $[x^* - r, x^* + r]$. Suppose that $x' \in [x^* - r, x^* + r]$ satisfies $f(x') = y$. Then we have $x' = \phi(x')$, and

$$|x - x'| = |\phi(x) - \phi(x')| \le \alpha|x - x'|. \tag{158}$$

As $\alpha < 1$, this implies that $x' = x$.

To conclude, so far, we have the following. Pick $0 < \alpha < 1$, and let $0 < r \leq \delta(\alpha)$ and $\varepsilon = \varepsilon(r, \alpha)$, where $\delta(\alpha)$ and $\varepsilon(r, \alpha)$ are as in Lemma 8.3. Then for any $y \in [y^* - \varepsilon, y^* + \varepsilon]$, there exists a unique $x \in [x^* - r, x^* + r]$ such that $f(x) = y$. Notice that this does not mean that each $x \in [x^* - r, x^* + r]$ is mapped to $y = f(x)$ in $[y^* - \varepsilon, y^* + \varepsilon]$, that is, we do not immediately get invertibility of f in $[x^* - r, x^* + r]$. This problem can be dealt with by choosing $r > 0$ sufficiently small, and by using continuity to guarantee that the interval $[x^* - r, x^* + r]$ is mapped into a region $[y^* - \varepsilon^*, y^* + \varepsilon^*]$ such that $f(x) = y$ can be solved for all $y \in [y^* - \varepsilon^*, y^* + \varepsilon^*]$. It is implemented in the proof of the following theorem.

Theorem 8.4 (Inverse function theorem, single variable version)**.** *Let $f : (a, b) \to \mathbb{R}$ be continuously differentiable in (a, b), and let $f'(x^*) \neq 0$ for some $x^* \in (a, b)$. Then there exists $r > 0$ such that f is invertible in $I = (x^* - r, x^* + r)$, and for $x \in I$, the inverse function is differentiable at $f(x)$ whenever $f'(x) \neq 0$, with*

$$(f^{-1})'(f(x)) = \frac{1}{f'(x)}. \tag{159}$$

In particular, f^{-1} and $(f^{-1})'$ are continuous at $f(x)$ with $x \in I$ whenever $f'(x) \neq 0$.

Proof. Let $0 < \alpha < 1$, and let $\varepsilon^* = \varepsilon(r^*, \alpha)$ with $r^* = \delta(\alpha)$. Then we know that for any $y \in [y^* - \varepsilon^*, y^* + \varepsilon^*]$, there exists a unique $x \in [x^* - r^*, x^* + r^*]$ such that $f(x) = y$. Now, by continuity, there exists $0 < r \leq r^*$ such that $|x - x^*| \leq r$ implies $|f(x) - y^*| \leq \varepsilon^*$. So for any $x \in [x^* - r, x^* + r]$, there exists a unique $z \in [x^* - r^*, x^* + r^*]$ satisfying $f(z) = f(x)$, and by uniqueness, we have $z = x$. This means that f is invertible in the interval $[x^* - r, x^* + r]$.

Let $I = (x^* - r, x^* + r)$, and let $\bar{x} \in I$ be such that $f'(\bar{x}) \neq 0$. The differentiability of f at $\bar{x}$ means that there exists $\varphi : (x^* - r, x^* + r) \to \mathbb{R}$, continuous at $\bar{x}$, such that

$$f(x) = f(\bar{x}) + \varphi(x)(x - \bar{x}), \qquad x \in I. \tag{160}$$

Since $\varphi(\bar{x}) = f'(\bar{x}) \neq 0$, by continuity, there is $\rho > 0$ such that $|\varphi(x)| \geq \frac{1}{2}|f'(\bar{x})|$ whenever $|x - \bar{x}| < \rho$. Thus for $x \in (\bar{x} - \rho, \bar{x} + \rho)$, we have

$$x - \bar{x} = \frac{f(x) - f(\bar{x})}{\varphi(x)}, \qquad \text{and} \qquad |x - \bar{x}| \leq \frac{2|f(x) - f(\bar{x})|}{|f'(\bar{x})|}. \tag{161}$$

With $y = f(x)$ and $\bar{y} = f(\bar{x})$, the latter can be written as

$$|f^{-1}(y) - f^{-1}(\bar{y})| \leq \frac{2|y - \bar{y}|}{|f'(\bar{x})|}, \tag{162}$$

which shows that f^{-1} is continuous at $\bar{y}$. The first equation in (161) becomes

$$f^{-1}(y) - f^{-1}(\bar{y}) = \psi(y)(y - \bar{y}), \qquad \text{where} \quad \psi(y) = \frac{1}{\varphi(f^{-1}(y))}. \tag{163}$$

As f^{-1} is continuous at $\bar{y}$ and φ is continuous at $\bar{x} = f^{-1}(\bar{y})$, the function ψ is continuous at $\bar{y}$, and hence f^{-1} is differentiable at $\bar{y}$, with

$$(f^{-1})'(\bar{y}) = \psi(\bar{y}) = \frac{1}{\varphi(f^{-1}(\bar{y}))} = \frac{1}{f'(f^{-1}(\bar{y}))} = \frac{1}{f'(\bar{x})}. \tag{164}$$

We conclude that (159) holds whenever $x \in I$ satisfies $f'(x) \neq 0$. Then $(f^{-1})' = \frac{1}{f' \circ f^{-1}}$ is continuous at $y = f(x)$ whenever $x \in I$ and $f'(x) \neq 0$, because both f' and f^{-1} would be continuous at x and $f(x)$, respectively. ∎

Exercise 8.5. The function $f(x) = x^3$ has the inverse $f^{-1}(y) = \sqrt[3]{y}$ for all $y \in \mathbb{R}$, but $f'(0) = 0$. How is this compatible with the inverse function theorem?

Exercise 8.6. Let $f : (a, b) \to (c, d)$ be a continuously differentiable function, whose inverse $f^{-1} : (c, d) \to (a, b)$ is also continuously differentiable. Show that $f'(x) \neq 0$ for all $x \in (a, b)$.

9. The inverse function theorem

The purpose of this section is to extend the inverse function theorem to n-dimensions. Let $Q = (a_1, b_1) \times \ldots \times (a_n, b_n) \subset \mathbb{R}^n$ be a rectangular domain, and let $f : Q \to \mathbb{R}^n$ be differentiable in Q. Then the derivative Df can be considered as a function sending Q to $\mathbb{R}^{n\times n}$, and we will assume that this function is continuous. In other words, we assume that $f : Q \to \mathbb{R}^n$ is *continuously differentiable in* Q. Furthermore, we consider some point $x^* \in Q$, and suppose that the matrix $Df(x^*) \in \mathbb{R}^{n\times n}$ is invertible. Recall that a matrix is invertible (or nonsingular) if and only if its determinant is nonzero. In light of the preceding section, given any $y \in \mathbb{R}^n$ that is close to $y^* = f(x^*)$, the plan is to design a sequence $x^{(i)} \in \mathbb{R}^n$, $i = 0, 1, \ldots$, whose limit would solve the equation $f(x) = y$. Thus, let $x^{(0)} = x^*$, and suppose that $x^{(i)} \in \mathbb{R}^n$ has been constructed. Then we can approximate $f(x)$ for $x \approx x_k$ by

$$f(x) \approx f(x^{(i)}) + Df(x^{(i)})(x - x^{(i)}), \tag{165}$$

and by continuity of Df, we have $Df(x^{(i)}) \approx Df(x^*)$ if $x^{(i)} \approx x^*$, which yields

$$f(x) \approx f(x^{(i)}) + Df(x^*)(x - x^{(i)}). \tag{166}$$

We expect that if we solve the equation

$$y = f(x^{(i)}) + Df(x^*)(x^{(i+1)} - x^{(i)}), \tag{167}$$

for the unknown $x^{(i+1)} \in \mathbb{R}^n$, then $x^{(i+1)}$ will be closer than $x^{(i)}$ to the hypothetical solution x of $f(x) = y$. Since $Df(x^*)$ is an invertible matrix, we can solve the latter equation as

$$x^{(i+1)} = x^{(i)} + [Df(x^*)]^{-1}(y - f(x^{(i)})), \tag{168}$$

and starting with $x^{(0)} = x^*$, we use it for $i = 0, 1, \ldots$ to construct the sequence $x^{(0)}, x^{(1)}, \ldots$. This is of course an extension of (142) to n-dimensions. To prove that the sequence $\{x^{(i)}\}$ converges, we begin by studying the map defined by the right hand side of (168).

We will be using the notation

$$\bar{Q}_r(y) = \{x \in \mathbb{R}^n : |x - y|_\infty \le r\} \equiv [y_1 - r, y_1 + r] \times \ldots \times [y_n - r, y_n + r], \tag{169}$$

for the n-dimensional cube centred at $y \in \mathbb{R}^n$, with the side length $2r$.

Lemma 9.1. *Let $Q \subset \mathbb{R}^n$ be a rectangular domain, and let $f : Q \to \mathbb{R}^n$ be continuously differentiable in Q. Suppose that $Df(x^*)$ is invertible for some $x^* \in Q$. Let $y \in \mathbb{R}^n$, and consider the map $\phi : Q \to \mathbb{R}^n$ defined by*

$$\phi(x) = x + [Df(x^*)]^{-1}(y - f(x)), \tag{170}$$

a) We have $\phi(x) = x$ if and only if $f(x) = y$.
b) For any $\alpha > 0$, there exists $\delta = \delta(\alpha) > 0$ such that

$$|\phi(x) - \phi(x')|_\infty \le \alpha |x - x'|_\infty, \qquad x, x' \in \bar{Q}_\delta(x^*). \tag{171}$$

c) Let $0 < \alpha < 1$, and let $\delta = \delta(\alpha)$ be as in b). Then for any $0 < r \le \delta$, there exists $\varepsilon = \varepsilon(\alpha, r) > 0$ such that

$$\phi(x) \in \bar{Q}_r(x^*) \qquad \textit{whenever} \quad x \in \bar{Q}_r(x^*) \quad \textit{and} \quad y \in \bar{Q}_\varepsilon(y^*), \tag{172}$$

where $y^* = f(x^*)$.

Proof. a) We have $\phi(x) - x = [Df(x^*)]^{-1}(y - f(x))$ or $Df(x^*)(\phi(x) - x) = y - f(x)$, so it is clear that $\phi(x) - x = 0$ and $y - f(x) = 0$ are equivalent.

b) If $x, x' \in Q$ are close to x^*, then we have

$$\begin{aligned} \phi(x) - \phi(x') &= x - x' + [Df(x^*)]^{-1}(f(x') - f(x)) \\ &= [Df(x^*)]^{-1}\left(f(x') - f(x) - Df(x^*)(x' - x)\right). \end{aligned} \tag{173}$$

Hence there is some $C > 0$ such that

$$|\phi(x) - \phi(x')|_\infty \le C \left|f(x') - f(x) - Df(x^*)(x' - x)\right|_\infty . \tag{174}$$

At this point, we invoke Lemma 9.2, proved below, to imply that the existence of $\delta = \delta(\alpha) > 0$ such that

$$|f(x') - f(x) - Df(x^*)(x' - x)|_\infty \le \frac{\alpha}{C}|x' - x|_\infty, \tag{175}$$

for all $x, x' \in \bar{Q}_\delta(x^*)$. Substituting this into the right hand side of (174), we get (171).

c) Now for $x \in Q$, we have

$$\begin{aligned}\phi(x) - x^* &= [Df(x^*)]^{-1}(y - f(x) + Df(x^*)(x - x^*)) \\ &= [Df(x^*)]^{-1}(f(x^*) - f(x) + Df(x)(x - x^*)) + [Df(x^*)]^{-1}(y - y^*).\end{aligned} \tag{176}$$

Let $0 < \alpha < 1$, and let $\delta = \delta(\alpha) > 0$ as in the previous paragraph. Then with $0 < r \le \delta$, for $x \in \bar{Q}_r(x^*)$, we have

$$\begin{aligned}|\phi(x) - x^*|_\infty &\le \alpha|x - x^*|_\infty + |[Df(x^*)]^{-1}(y - y^*)|_\infty \\ &\le \alpha r + C|y - y^*|_\infty.\end{aligned} \tag{177}$$

This means that if $|y - y^*| \le (1 - \alpha)r/C$, then $\phi(x) \in \bar{Q}_r(x^*)$ whenever $x \in \bar{Q}_r(x^*)$. □

The following result has been used in the preceding proof.

Lemma 9.2. *Let $Q \subset \mathbb{R}^n$ be a rectangular domain, and let $f : Q \to \mathbb{R}^n$ be continuously differentiable in Q. Suppose that $Df(x^*)$ is invertible for some $x^* \in Q$. Then given any $\beta > 0$, there exists $\delta = \delta(\beta) > 0$ such that*

$$|f(x') - f(x) - Df(x^*)(x' - x)|_\infty \le \beta|x' - x|_\infty, \qquad x, x' \in \bar{Q}_\delta(x^*). \tag{178}$$

Proof. Let $x, x' \in Q$, and let $g(t) = f(x + tV)$, where $V = x' - x$. Then we have

$$f(x') - f(x) = g(1) - g(0), \tag{179}$$

and

$$g(t) = D_V f(x + tV) = Df(x + tV)V. \tag{180}$$

By the mean value theorem, for each k, there exists $0 < t_k < 1$ such that $g_k(1)-g_k(0) = g_k'(t_k)$, that is,

$$f_k(x) - f_k(x') = D_V f_k(x + t_kV) = Df_k(x + t_kV)V = Df_k(x + t_kV)(x' - x) \tag{181}$$

This implies that

$$f_k(x') - f_k(x) - Df_k(x^*)(x' - x) = (Df_k(x + t_kV) - Df_k(x^*))(x' - x), \tag{182}$$

and so

$$|f_k(x') - f_k(x) - Df_k(x^*)(x' - x)| \le \sum_{i=1}^n |\partial_i f_k(x + t_kV) - \partial_i f_k(x^*)||x_i' - x_i|. \tag{183}$$

Note that $\partial_i f_k$ is simply an entry in the Jacobian matrix (or the derivative) Df, and by continuity of Df, the quantity $\partial_i f_k(x + t_kV) - \partial_i f_k(x^*)$ would be arbitrarily small if x and x' are close to x^*. More specifically, for any $\beta > 0$, there exists $\delta = \delta(\beta) > 0$ such that $|\partial_i f_k(z) - \partial_i f_k(x^*)| < \beta/n$ whenever $z \in \bar{Q}_\delta(x^*)$. Then we have $x + t_kV \in \bar{Q}_\delta(x^*)$ for $x, x' \in \bar{Q}_\delta(x^*)$, and thus

$$|f_k(x') - f_k(x) - Df_k(x^*)(x' - x)| \le n\frac{\beta}{n} \max_i |x_i' - x_i| \le \beta|x' - x|_\infty, \tag{184}$$

for $x, x' \in \bar{Q}_\delta(x^*)$. □

Next, we show that under certain conditions, the sequence $\{x^{(i)}\}$ converges, and the limit is the unique solution of the equation $f(x) = y$.

Lemma 9.3. *Let $Q \subset \mathbb{R}^n$ be a rectangular domain, and let $f : Q \to \mathbb{R}^n$ be continuously differentiable in Q. Suppose that $Df(x^*)$ is invertible for some $x^* \in Q$. Then there exists $\delta > 0$ with the following property. For any $0 < r \leq \delta$, there exists $\varepsilon = \varepsilon(r) > 0$, such that the equation $f(x) = y$ has a unique solution x in $\bar{Q}_r(x^*)$ whenever $y \in \bar{Q}_\varepsilon(y^*)$, where $y^* = f(x^*)$.*

Proof. Let $0 < \alpha < 1$, and let $\delta = \delta(\alpha) > 0$ be as in Lemma 9.1b). Furthermore, given an arbitrary $0 < r \leq \delta$, let $\varepsilon = \varepsilon(\alpha, r)$ be as in Lemma 9.1c). Next, we let $y \in \bar{Q}_\varepsilon(y^*)$, and define the sequence $\{x^{(i)}\} \subset \bar{Q}_r(x^*)$ by $x^{(0)} = x^*$, and $x^{(i+1)} = \phi(x^{(i)})$ for $i = 0, 1, \ldots$. Then we have

$$|x^{(i+1)} - x^{(i)}|_\infty = |\phi(x^{(i)}) - \phi(x^{(i-1)})|_\infty \leq \alpha |x^{(i)} - x^{(i-1)}|_\infty \leq \ldots \leq \alpha^i |x^{(1)} - x^{(0)}|_\infty, \quad (185)$$

and hence

$$\begin{aligned} |x^{(i+m)} - x^{(i)}|_\infty &\leq |x^{(i+m)} - x^{(i+m-1)}|_\infty + \ldots + |x^{(i+1)} - x^{(i)}|_\infty \\ &\leq (\alpha^{i+m-1} + \ldots + \alpha^i)|x^{(1)} - x^{(0)}|_\infty \\ &\leq \alpha^i (1 + \alpha + \alpha^2 + \ldots)|x^{(1)} - x^{(0)}|_\infty = \frac{\alpha^i}{1-\alpha}|x^{(1)} - x^{(0)}|_\infty, \end{aligned} \quad (186)$$

showing that $|x^{(m)} - x^{(i)}|_\infty \to 0$ as $\min\{i, m\} \to \infty$. This implies that $|x_k^{(m)} - x_k^{(i)}|_\infty \to 0$ as $\min\{i, m\} \to \infty$ for each $k \in \{1, \ldots, n\}$, where $x_k^{(i)}$ denotes the k-th component of $x^{(i)} \in \mathbb{R}^n$. In other words, for each $k \in \{1, \ldots, n\}$, the scalar sequence $x_k^{(0)}, x_k^{(1)}, \ldots$ is Cauchy, and hence there exists $x_k \in [x_k^* - r, x_k^* + r]$ such that $x_k^{(i)} \to x_k$ as $i \to \infty$. If we collect the components x_k into one vector $x = (x_1, \ldots, x_n)$, then we have $x \in \bar{Q}_r(x^*)$ and $x^{(i)} \to x$ as $i \to \infty$.

Now we want to show that $f(x) = y$, or equivalently, that $x = \phi(x)$. For any i, we have

$$\begin{aligned} |x - \phi(x)|_\infty = |x - x^{(i+1)} + \phi(x^{(i)}) - \phi(x)|_\infty &\leq |x - x^{(i+1)}| + |\phi(x^{(i)}) - \phi(x)|_\infty \\ \leq |x - x^{(i+1)}|_\infty + \alpha |x^{(i)} - x|_\infty, \end{aligned} \quad (187)$$

and the right hand side tends to 0 as $i \to \infty$. That is, we have $|x - \phi(x)|_\infty \leq e$ for any $e > 0$, which means that $|x - \phi(x)|_\infty = 0$, and hence $x = \phi(x)$.

Finally, we need to show that x is the only solution of $f(x) = y$ in $\bar{Q}_r(x^*)$. Suppose that $x' \in \bar{Q}_r(x^*)$ satisfies $f(x') = y$. Then we have $x' = \phi(x')$, and so

$$|x - x'|_\infty = |\phi(x) - \phi(x')|_\infty \leq \alpha |x - x'|_\infty. \quad (188)$$

As $\alpha < 1$, this implies that $x' = x$. □

Before stating the final theorem, let us prove one more preliminary lemma.

Lemma 9.4. *Let $K \subset \mathbb{R}^n$ be a set, and let $G : K \to \mathbb{R}^{m \times m}$ be a matrix-valued function. Assume that G is continuous at $y \in K$, and that $G(y)$ is an invertible matrix. Then there exists $r > 0$ such that $G(x)$ is invertible whenever $x \in \bar{Q}_r(y)$. Moreover, $[G(x)]^{-1}$ is continuous at y as a matrix-valued function of x.*

Proof. Recall that a matrix is invertible if and only if its determinant is nonzero, and the determinant of a matrix is a polynomial of the matrix entries. In particular, the determinant is continuous in $\mathbb{R}^{m \times m}$ as a function $\det : \mathbb{R}^{m \times m} \to \mathbb{R}$. This means that the function $g : K \to \mathbb{R}$, defined by $g(x) = \det G(x)$ for $x \in K$, is continuous in K. Since $G(y)$ is invertible, we have $g(y) \neq 0$, and by continuity of g, there exists $r > 0$ such that $x \in \bar{Q}_r(y)$ implies $|g(x) - g(y)| < \frac{1}{2}|g(y)|$, and hence $|g(x)| > \frac{1}{2}|g(y)| > 0$. This means that $G(x)$ is invertible for all $x \in \bar{Q}_r(y)$.

Introducing the notations $A = [G(y)]^{-1}$, $Y = [G(x)]^{-1} - A$, and $X = G(x) - G(y)$, where $x \in \bar{Q}_r(y)$, the identity $[G(x)]^{-1}G(x) = I$ is written as

$$I = (A + Y)(G(y) + X) = AG(y) + YG(y) + AX + YX. \tag{189}$$

Taking into account that $AG(y) = I$, we get

$$YG(y) = -AX - YX, \tag{190}$$

and multiplying both sides by A gives

$$Y = -AXA - YXA. \tag{191}$$

For $B, C \in \mathbb{R}^{n\times n}$, and $E = BC$, we have

$$|E_{ik}| \leq \sum_{j=1}^{n} |B_{ij}C_{jk}| \leq n|B|_\infty |C|_\infty, \tag{192}$$

which yields $|E|_\infty \leq n|B|_\infty|C|_\infty$. We apply this estimate to (191), and infer

$$|Y|_\infty \leq n^2|A|_\infty^2|X|_\infty + n^2|A|_\infty|X|_\infty|Y|_\infty. \tag{193}$$

Recall that $X = G(x) - G(y)$, and we choose $\rho > 0$ so small that $n^2|A|_\infty|G(x) - G(y)|_\infty \leq \frac{1}{2}$ for all $x \in \bar{Q}_\rho(y)$. Then for $x \in \bar{Q}_\rho(y)$, we have

$$|Y|_\infty \leq n^2|A|_\infty^2|X|_\infty + \tfrac{1}{2}|Y|_\infty. \tag{194}$$

Subtracting $\frac{1}{2}|Y|_\infty$ from both sides, and multiplying both sides by 2, we get

$$|Y|_\infty \leq 2n^2|A|_\infty^2|X|_\infty \qquad \text{for} \quad x \in \bar{Q}_\rho(y). \tag{195}$$

Since $A = [G(y)]^{-1}$ is a fixed matrix, and $|X|_\infty$ can be made arbitrarily small by choosing $|x - y|_\infty$ small, we conclude that $[G(x)]^{-1}$ is continuous at y as a function of x. □

We are now ready to state and prove the main result of this section. Introduce the notation

$$Q_r(y) = \{x \in \mathbb{R}^n : |x - y|_\infty < r\} \equiv (y_1 - r, y_1 + r) \times \ldots \times (y_n - r, y_n + r), \tag{196}$$

for the n-dimensional cube centred at $y \in \mathbb{R}^n$, with the side length $2r$. This is called an *open* cube, as opposed the the *closed* cube $\bar{Q}_r(y)$. Note that we always have $Q_r(y) \subset \bar{Q}_r(y)$.

Theorem 9.5 (Inverse function theorem)**.** *Let $Q \subset \mathbb{R}^n$ be a rectangular domain, and let $f : Q \to \mathbb{R}^n$ be continuously differentiable in Q. Suppose that $Df(x^*)$ is invertible for some $x^* \in Q$. Then there exists $r > 0$ such that f is invertible in $Q_r(x^*)$, and the inverse function is differentiable in $f(Q_r(x^*))$, with*

$$Df^{-1}(f(x)) = (Df(x))^{-1}, \qquad x \in Q_r(x^*). \tag{197}$$

In particular, f^{-1} and Df^{-1} are continuous in $f(Q_r(x^*))$, and moreover, there is $\varepsilon > 0$ such that $Q_\varepsilon(y^*) \subset f(Q_r(x^*))$.

Proof. Let $\delta > 0$ be as in Lemma 9.3, and let $\varepsilon^* = \varepsilon(\delta)$. Then Lemma 9.3 guarantees that for any $y \in \bar{Q}_{\varepsilon^*}(y^*)$, there exists a unique $x \in \bar{Q}_\delta(x^*)$ such that $f(x) = y$. By continuity of f, there exists $0 < r \leq \delta$ such that $x \in \bar{Q}_r(x^*)$ implies $f(x) \in \bar{Q}_{\varepsilon^*}(y^*)$. So for any $x \in \bar{Q}_r(x^*)$, there exists a unique $z \in \bar{Q}_\delta(x^*)$ satisfying $f(z) = f(x)$, and by uniqueness, we have $z = x$. This means that f is invertible in $\bar{Q}_r(x^*)$.

Since $Df(x^*)$ is invertible and Df is continuous, by Lemma 9.4, choosing $r > 0$ smaller if necessary, we can assume that $Df(x)$ is invertible for all $x \in \bar{Q}_r(x^*)$.

Let $\bar{x} \in Q_r(x^*)$. Then by differentiability of f, there exists $G : Q_r(x^*) \to \mathbb{R}^{n\times n}$, continuous at $\bar{x}$, such that

$$f(x) = f(\bar{x}) + G(x)(x - \bar{x}), \qquad x \in Q_r(x^*). \tag{198}$$

Since $G(\bar{x}) = Df(\bar{x})$ is invertible, by Lemma 9.4, there is $\rho > 0$ such that $A(x) = [G(x)]^{-1}$ exists whenever $x \in Q_\rho(\bar{x})$, and $x \mapsto A(x)$ is continuous at $\bar{x}$. Thus for $x \in Q_\rho(\bar{x})$, we have

$$x - \bar{x} = A(x)\,(f(x) - f(\bar{x}))\,, \qquad \text{and so} \qquad |x - \bar{x}|_\infty \leq n|A(x)|_\infty |f(x) - f(\bar{x})|_\infty. \tag{199}$$

The norm $|A(x)|_\infty$ is a continuous function of x at $\bar{x}$, and hence there is some constant $C > 0$ such that $n|A(x)|_\infty \leq C$ for all $x \in Q_\rho(\bar{x})$, with $\rho > 0$ possibly smaller than its original value.

In any case, with $y = f(x)$ and $\bar{y} = f(\bar{x})$, and for some constant C, we have

$$|f^{-1}(y) - f^{-1}(\bar{y})|_\infty \leq C|y - \bar{y}|_\infty, \tag{200}$$

which shows that f^{-1} is continuous at $\bar{y}$. The first equation in (199) becomes

$$f^{-1}(y) - f^{-1}(\bar{y}) = \Psi(y)(y - \bar{y}), \qquad \text{where} \quad \Psi(y) = A(f^{-1}(y)). \tag{201}$$

As f^{-1} is continuous at $\bar{y}$ and A is continuous at $\bar{x} = f^{-1}(\bar{y})$, the function Ψ is continuous at $\bar{y}$, and hence f^{-1} is differentiable at $\bar{y}$, with

$$Df^{-1}(\bar{y}) = \Psi(\bar{y}) = A(f^{-1}(\bar{y})) = [Df(f^{-1}(\bar{y}))]^{-1} = [Df(\bar{x})]^{-1}. \tag{202}$$

This establishes (197). Then $Df^{-1} = [Df \circ f^{-1}]^{-1}$ is continuous because both f' and f^{-1} are continuous. □

Example 9.6. Consider the map $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$, defined by $\Psi(r, \phi) = (r\cos\phi, r\sin\phi)$. With $x = x(r, \phi) = r\cos\phi$ and $y = y(r, \phi) = r\sin\phi$ denoting the components of Ψ, the Jacobian of Ψ and its determinant are given by

$$J = \begin{pmatrix} \partial_r x & \partial_\phi x \\ \partial_r y & \partial_\phi y \end{pmatrix} = \begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix}, \qquad \text{and} \qquad \det J = r. \tag{203}$$

Since J is a continuous function of $(r, \phi) \in \mathbb{R}^2$, the map Ψ is differentiable in $\mathbb{R}^2$, with $D\Psi = J$. Moreover, $D\Psi(r, \phi)$ is invertible whenever $r \neq 0$, and

$$(D\Psi)^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\phi & r\sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}. \tag{204}$$

By the inverse function theorem, for any $(r^*, \phi^*) \in \mathbb{R}^2$ with $r^* \neq 0$, there exists $\delta > 0$ such that Ψ is invertible in $(r^* - \delta, r^* + \delta) \times (\phi^* - \delta, \phi^* + \delta)$, with

$$D\Psi^{-1}(x, y) = \begin{pmatrix} \partial_x r & \partial_y r \\ \partial_x \phi & \partial_y \phi \end{pmatrix} = \frac{1}{r}\begin{pmatrix} r\cos\phi & r\sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}, \tag{205}$$

where $r = r(x, y)$ and $\phi = \phi(x, y)$ are now understood to be the components of Ψ^{-1}. Note that Ψ^{-1} is guaranteed to satisfy $\Psi^{-1}(\Psi(r, \phi)) = (r, \phi)$ for all $(r, \phi) \in (r^* - \delta, r^* + \delta) \times (\phi^* - \delta, \phi^* + \delta)$, and nothing more, so that we would have a potentially different inverse function Ψ^{-1} to Ψ if we change the centre $(r^*, \phi^*) \in \mathbb{R}^2$ and apply the inverse function theorem again. In practice, it does not cause much trouble because we usually work in one such region at a time.

Let $u : \mathbb{R}^2 \to \mathbb{R}$ be a twice continuously differentiable function, and let $v = u \circ \Psi$. We will think of u as a function of $(x, y) \in \mathbb{R}^2$, and v as a function of $(r, \phi) \in \mathbb{R}^2$. The chain rule gives

$$Dv(r, \phi) = Du(\Psi(r, \phi))D\Psi(r, \phi), \tag{206}$$

where Dv and Du should be considered as row vectors. In components, this is

$$\begin{aligned} \partial_r v &= \partial_x u\,\partial_r x + \partial_y u\,\partial_r y = \cos\phi\,\partial_x u + \sin\phi\,\partial_y u, \\ \partial_\phi v &= \partial_x u\,\partial_\phi x + \partial_y u\,\partial_\phi y = -r\sin\phi\,\partial_x u + r\cos\phi\,\partial_y u. \end{aligned} \tag{207}$$

Since u is twice continuously differentiable, the functions in the right hand side are continuously differentiable, and hence we can apply the chain rule again to infer that v is twice continuously differentiable. For instance, we can compute

$$\begin{aligned}
\partial_r^2 v &= \cos\phi\,\partial_x(\cos\phi\,\partial_x u + \sin\phi\,\partial_y u) + \sin\phi\,\partial_y(\cos\phi\,\partial_x u + \sin\phi\,\partial_y u) \\
&= -\cos\phi\sin\phi\,\partial_x\phi\,\partial_x u + \cos^2\phi\,\partial_x^2 u + \cos^2\phi\,\partial_x\phi\,\partial_y u + \sin\phi\cos\phi\,\partial_x\partial_y u \\
&\quad - \sin^2\phi\,\partial_y\phi\,\partial_x u + \sin\phi\cos\phi\,\partial_y\partial_x u + \sin\phi\cos\phi\,\partial_y\phi\,\partial_y u + \sin^2\phi\,\partial_y^2 u \\
&= \frac{\cos\phi\sin^2\phi}{r}\,\partial_x u + \cos^2\phi\,\partial_x^2 u - \frac{\cos^2\phi\sin\phi}{r}\,\partial_y u + \sin\phi\cos\phi\,\partial_x\partial_y u \\
&\quad - \frac{\sin^2\phi\cos\phi}{r}\,\partial_x u + \sin\phi\cos\phi\,\partial_y\partial_x u + \frac{\sin\phi\cos^2\phi}{r}\,\partial_y u + \sin^2\phi\,\partial_y^2 u \\
&= \cos^2\phi\,\partial_x^2 u + 2\sin\phi\cos\phi\,\partial_x\partial_y u + \sin^2\phi\,\partial_y^2 u,
\end{aligned} \tag{208}$$

where we have used the expressions for $\partial_x\phi$ and $\partial_y\phi$ from (205).

Now we consider the inverse Ψ^{-1} of Ψ in some region given by the inverse function theorem. In what follows, all points (x, y) will be assumed to be in the region where Ψ is invertible, and all (r, ϕ) will be assumed to be in the region where $\Psi^{-1}(r, \phi)$ makes sense. So, by the chain rule, we have

$$Du(x, y) = Dv(\Psi^{-1}(x, y))D\Psi^{-1}(x, y), \tag{209}$$

which is written in components as

$$\begin{aligned}
\partial_x u &= \partial_r v\,\partial_x r + \partial_\phi v\,\partial_x\phi = \cos\phi\,\partial_r v - \frac{\sin\phi}{r}\,\partial_\phi v, \\
\partial_y u &= \partial_r v\,\partial_y r + \partial_\phi v\,\partial_y\phi = \sin\phi\,\partial_r v + \frac{\cos\phi}{r}\,\partial_\phi v.
\end{aligned} \tag{210}$$

Note that in the right hand side we have functions of (r, ϕ), and they should be evaluated at $(r, \phi) = \Psi^{-1}(x, y)$. For instance, we have $\partial_x u = w \circ \Psi^{-1}$, where

$$w(r, \phi) = \cos\phi\,\partial_r v(r, \phi) - \frac{\sin\phi}{r}\,\partial_\phi v(r, \phi). \tag{211}$$

We see that what w is to $\partial_x u$ is exactly what v is to u. Hence we can apply the first equality of (210) to the pair w and $\partial_x u$, and infer

$$\begin{aligned}
\partial_x^2 u = \partial_x(\partial_x u) &= \cos\phi\,\partial_r w - \frac{\sin\phi}{r}\,\partial_\phi w \\
&= \cos\phi\left(\cos\phi\,\partial_r^2 v + \frac{\sin\phi}{r^2}\,\partial_\phi v - \frac{\sin\phi}{r}\,\partial_r\partial_\phi v\right) \\
&\quad - \frac{\sin\phi}{r}\left(-\sin\phi\,\partial_r v + \cos\phi\,\partial_\phi\partial_r v - \frac{\cos\phi}{r}\,\partial_\phi v - \frac{\sin\phi}{r}\,\partial_\phi^2 v\right) \\
&= \cos^2\phi\,\partial_r^2 v - \frac{2\sin\phi\cos\phi}{r}\,\partial_\phi\partial_r v + \frac{\sin^2\phi}{r^2}\,\partial_\phi^2 v + \frac{\sin^2\phi}{r}\,\partial_r v + \frac{2\sin\phi\cos\phi}{r^2}\,\partial_\phi v.
\end{aligned} \tag{212}$$

A similar computation gives

$$\partial_y^2 u = \sin^2\phi\,\partial_r^2 v + \frac{2\sin\phi\cos\phi}{r}\,\partial_\phi\partial_r v + \frac{\cos^2\phi}{r^2}\,\partial_\phi^2 v + \frac{\cos^2\phi}{r}\,\partial_r v - \frac{2\sin\phi\cos\phi}{r^2}\,\partial_\phi v, \tag{213}$$

and summing the last two expressions, we get

$$\partial_x^2 u + \partial_y^2 u = \partial_r^2 v + \frac{1}{r^2}\,\partial_\phi^2 v + \frac{1}{r}\,\partial_r v. \tag{214}$$

This is the well known expression for the *Laplacian* $\Delta u = \partial_x^2 u + \partial_y^2 u$ in polar coordinates.

Exercise 9.7. In the context of the preceding example, compute $\partial_\phi^2 v$, $\partial_\phi\partial_r v$, and $\partial_x\partial_y u$.