Lecture Notes
Analysis II

For Engineering Students

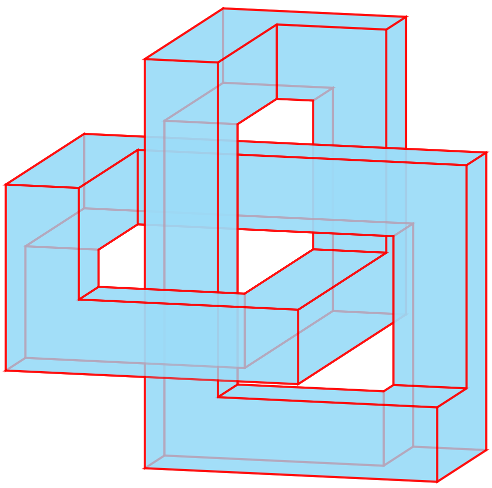
Spring Semester 2025

Contents

Chapter 1

The Euclidean space $\mathbb{R}^n$

In Analysis 1 you have learned the fundamental concepts of differential and integral calculus of real-valued functions in one real variable, known as *Single Variable Calculus*. However, real-life phenomena often depend on a multitude of factors and it requires more than just one variable to properly model such situations. This leads to the study of the theory of differentiation and integration of functions in several variables, called *Multivariable Calculus*. The mathematical stage on which the study of functions in several variables unfolds is the n-dimensional Euclidean space $\mathbb{R}^n$.



Before defining the n-dimensional Euclidean space and its intrinsic topology, let us recall some basic notions commonly used in analysis and calculus.

- $\mathbb{N}$ the *natural numbers* $\{1, 2, 3, 4, \ldots\}$,
- $\mathbb{Z}$ the *integers*, i.e., signed whole numbers $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$,
- $\mathbb{Q}$ the *rational numbers* $\frac{a}{b}$ with $a \in \mathbb{Z}$ and $b \in \mathbb{N}$,
- $\mathbb{R}$ the *real numbers*,
- $\mathbb{C}$ the *complex numbers*,

An *open interval* is an interval that does not include its boundary points and is denoted by parentheses. The open intervals are thus one of the forms

$$\begin{aligned}(a,b) &= \{x\in\mathbb{R} : a<x<b\},\\ (-\infty,b) &= \{x\in\mathbb{R} : x<b\},\\ (a,+\infty) &= \{x\in\mathbb{R} : a<x\},\\ (-\infty,+\infty) &= \mathbb{R},\end{aligned}$$

where a and b are real numbers with $a \leqslant b$. The interval $(a,a)=\emptyset$ is the empty set, a degenerate interval. Open intervals are *open sets* in the topology of $\mathbb{R}$.

A *closed interval* is an interval that includes all its boundary points and is denoted by square brackets. Closed intervals take the form

$$\begin{aligned}[a,b] &= \{x\in\mathbb{R} : a\leqslant x\leqslant b\},\\ (-\infty,b] &= \{x\in\mathbb{R} : x\leqslant b\},\\ [a,+\infty) &= \{x\in\mathbb{R} : a\leqslant x\},\\ (-\infty,+\infty) &= \mathbb{R},\end{aligned}$$

Closed intervals are *closed sets* in the topology of $\mathbb{R}$. Note that the interval $\mathbb{R} = (-\infty,+\infty)$ is both open and closed at the same time.

A *half-open interval* is a finite interval that includes one endpoint but not the other. It can be left-open or right-open, depending on which endpoint is excluded:

$$\begin{aligned}(a,b] &= \{x\in\mathbb{R} : a<x\leqslant b\},\\ [a,b) &= \{x\in\mathbb{R} : a\leqslant x<b\},\end{aligned}$$

Note that half-open intervals are neither open nor closed sets in the topology of $\mathbb{R}$.

Intervals of the form $[a,b]$, $[a,b)$, $(a,b]$, (a,b) for $a,b\in\mathbb{R}$ with $a\leqslant b$ are called *bounded intervals*, whereas intervals like $(-\infty,b]$, $(-\infty,b)$, $[a,+\infty)$, and $(a,+\infty)$ are *unbounded intervals*.

1.1 The vector space $\mathbb{R}^n$

Given a positive integer n, the set $\mathbb{R}^n$ is defined as the set of all ordered n-tuples $(x_1,\ldots,x_n)$ of real numbers. It is called the *standard Euclidean space of dimension* n, or simply the n*-dimensional Euclidean space*.

We can represent an element of $\mathbb{R}^n$ either as an n-tuple, which is the same as a row vector with n entries,

$$\mathbf{x} = (x_1,\ldots,x_n)$$

or as a column vector with n entries

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

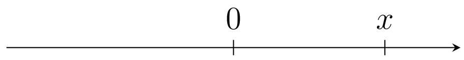
Both representations are common and widely used in the literature. We will generally use column vectors to denote elements of $\mathbb{R}^n$ in calculations, and row vectors to denote elements of $\mathbb{R}^n$ as input parameters of functions defined on $\mathbb{R}^n$.

There are also different ways in which elements in $\mathbb{R}^n$ are denoted, the three most common are

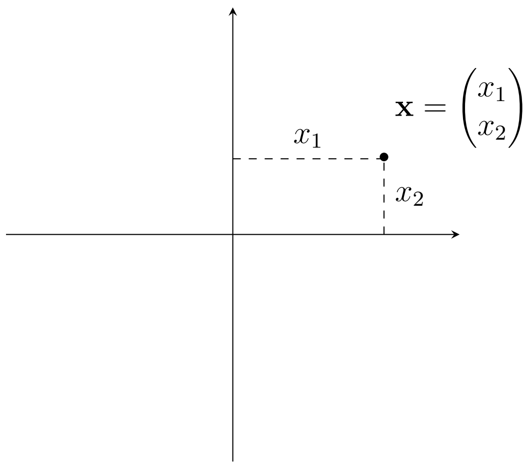
$$x, \qquad \mathbf{x}, \qquad \text{and} \qquad \vec{x}.$$

In this text, we will predominantly use x for elements in $\mathbb{R}$ and $\mathbf{x}$ for elements in $\mathbb{R}^n$ for $n \geqslant 2$.

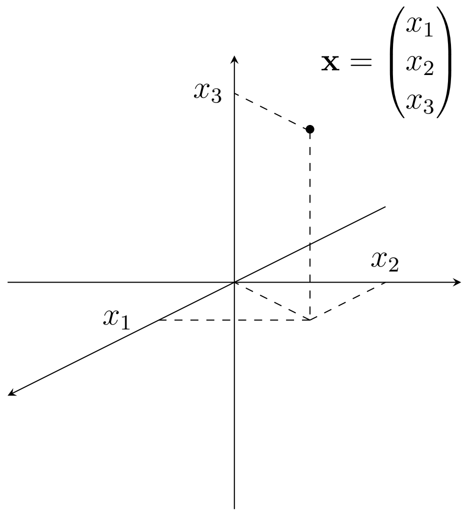
If $n = 1$ then $\mathbb{R}^1 = \mathbb{R}$ corresponds to the real line.



If $n = 2$ then $\mathbb{R}^2$ corresponds to the 2-dimensional plane. A point in $\mathbb{R}^2$ is usually written as either (x, y) or $\mathbf{x} = (x_1, x_2)^\top$.



If $n = 3$ then $\mathbb{R}^3$ corresponds to the 3-dimensional space. A point in $\mathbb{R}^3$ is usually written as eitehr (x, y, z) or $\mathbf{x} = (x_1, x_2, x_3)^\top$.



The set $\mathbb{R}^n$ is an n-dimensional inner product vector space over the real numbers. This means it is closed under addition, scalar multiplication, and endowed with an inner product called the scalar product. The addition on $\mathbb{R}^n$ is defined coordinate wise by

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

The multiplication of an element $\mathbf{x} \in \mathbb{R}^n$ by a scalar $\lambda \in \mathbb{R}$ is defined as

$$\lambda \mathbf{x} = \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

The way in which addition and multiplication on $\mathbb{R}^n$ interact is described by the distributive law, which asserts that

$$\lambda(\mathbf{x} + \mathbf{y}) = \lambda \mathbf{x} + \lambda \mathbf{y}. \qquad \text{(Distributive Law)}$$

The vector space $\mathbb{R}^n$ is also equipped with a *scalar product* $\langle .,. \rangle \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{n} x_k y_k. \tag{1.1}$$

The scalar product satisfies the three following properties:

1. **Positive-definiteness:** $\langle \mathbf{x}, \mathbf{x} \rangle \geqslant 0$ for all $\mathbf{x} \in \mathbb{R}^n$, with equality only for $\mathbf{x} = \mathbf{0}$.
2. **Symmetry:** $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
3. **Bilinearity:** $\langle \alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$.

In linear algebra, a vector $\mathbf{x}$ is also an $n \times 1$ matrix. Its transpose, written $\mathbf{x}^\top = (x_1, \ldots, x_n)$, is therefore a $1 \times n$ matrix, and we can interpret the scalar product of two vectors $\mathbf{x}, \mathbf{y}$ as the matrix product of $\mathbf{x}^\top$ and $\mathbf{y}$:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \cdot \mathbf{y} = (x_1, \ldots, x_n) \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

1.2 The Euclidean distance on $\mathbb{R}^n$

To be able to extend the analytical methods presented in Analysis 1 to the space $\mathbb{R}^n$, it is important to endow $\mathbb{R}^n$ with a topological structure. On $\mathbb{R}$ we have used the absolute value to define a distance $d(x, y) = |x - y|$, which was then used to define notions such as convergence and continuity in $\mathbb{R}$. We seek to generalize the absolute value and the distance to the space $\mathbb{R}^n$. To do so, we will introduce the concepts of a norm and a metric.

Definition 1.1 (The Euclidean norm on $\mathbb{R}^n$)**.** The *Euclidean norm* on $\mathbb{R}^n$ is the function $\|.\|_2 : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \left(\sum_{k=1}^{n} x_k^2 \right)^{\frac{1}{2}}. \tag{1.2}$$

It measures the distance of the point $\mathbf{x}$ to the origin $\mathbf{0} = (0, \dots, 0)$.

Observe that in one dimension, the Euclidean norm of a real number is the same as the absolute value of that number. In general, the Euclidean norm satisfies the following properties:

1. **Non-negativity:** $\|\mathbf{x}\|_2 \geqslant 0$ for all $\mathbf{x} \in \mathbb{R}^n$, with equality if and only if $\mathbf{x} = \mathbf{0}$.
2. **Homogeneity:** $\|\lambda \cdot \mathbf{x}\|_2 = |\lambda| \cdot \|\mathbf{x}\|_2$ for all $\lambda \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$.
3. **Triangle inequality:** $\|\mathbf{x} + \mathbf{y}\|_2 \leqslant \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

One of the most important properties of the scalar product is the *Cauchy-Schwarz inequality*, which says that

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leqslant \|\mathbf{x}\|_2 \, \|\mathbf{y}\|_2 \tag{Cauchy-Schwarz}$$

The Euclidean norm $\|\mathbf{x}\|_2$ also corresponds to the length of a vector $\mathbf{x}$. The scalar product $\langle \mathbf{x}, \mathbf{y} \rangle$ measures the angle between the two vectors $\mathbf{x}$ and $\mathbf{y}$: if we designate θ as the angle between $\mathbf{x}$ and $\mathbf{y}$, then

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \cos\theta. \tag{Angle Formula}$$

In particular if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal vectors, i.e., $\theta = \pm\pi/2$, then $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. As a consequence, we obtain the famous *Pythagorean theorem*, which says that if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal then

$$\|\mathbf{x} + \mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2. \tag{Pythagoras}$$

With the help of the Euclidean norm we can define a metric on $\mathbb{R}^n$ called the Euclidean distance.

Definition 1.2 (The Euclidean distance on $\mathbb{R}^n$)**.** The *Euclidean distance* on $\mathbb{R}^n$ is the function $d(.,.) \colon \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ given by

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}. \tag{1.3}$$

The Euclidean distance captures the natural distance between two points in $\mathbb{R}^n$. It satisfies the following three properties:

1. **Non-negativity:** $d(\mathbf{x}, \mathbf{y}) \geqslant 0$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, with equality only when $\mathbf{x} = \mathbf{y}$.
2. **Symmetry:** $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.
3. **Triangle inequality:** $d(\mathbf{x}, \mathbf{y}) \leqslant d(\mathbf{x}, \mathbf{z}) + d(\mathbf{y}, \mathbf{z})$.

1.3 The topology on $\mathbb{R}^n$

The Euclidean distance $d(\mathbf{x}, \mathbf{y})$ induces a topology on $\mathbb{R}^n$ which underpins all analytical considerations on $\mathbb{R}^n$. In particular, notions such as continuity, convergence, differentiablility and integrability are all defined in terms of this topology. The building blocks of the topology on $\mathbb{R}^n$ are the so-called open balls.

Definition 1.3 (Open Ball)**.** Let $\mathbf{a} \in \mathbb{R}^n$ and $r > 0$. The set

$$B(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{a}) < r\}$$

is called the *open ball* of radius r centered at $\mathbf{a}$.

Open balls are the mathematical conceptualization of "nearness" and an important use of open balls is to topologically distinguish distinct points: if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{x} \neq \mathbf{y}$ then we can find a sufficiently small open ball centered at $\mathbf{x}$ and another sufficiently small open ball centered at $\mathbf{y}$ such that these two balls don't touch.

Open balls are instances of open sets. An open set is a set with the property that if $\mathbf{x}$ is a point in the set then all points that are sufficiently near to $\mathbf{x}$ also belong to the set. The mathematically precise definition is as follows:

Definition 1.4 (Open set)**.** A subset $U \subseteq \mathbb{R}^n$ is *open* if for any point $\mathbf{x} \in U$ there exists $\varepsilon > 0$ such that the open ball $B(\mathbf{x}, \varepsilon)$ is contained in U.

The empty set $\emptyset$ and the space $\mathbb{R}^n$ are open. Also, as was already mentioned, any open ball $B(\mathbf{a}, r)$ is an open set.

Example 1.1 (Open Sets in $\mathbb{R}^n$)**.**

1. If $a < b$ are real numbers then the interval

$$(a, b) = \{x \in \mathbb{R} : a < x < b\}$$

   is an open set. Indeed, if $x \in (a, b)$, simply take $r = \min\{x - a, b - x\}$. Both these numbers are strictly positive, since $a < x < b$, and so is their minimum. Then the "1-dimensional ball" $B(x, r) = \{y \in \mathbb{R} : |x - y| < r\}$ is a subset of (a, b). This proves that (a, b) is an open set.
2. The infinite intervals (a, ∞) and $(-\infty, b)$ are also open but the intervals

$$(a, b] = \{x \in \mathbb{R} : a < x \leqslant b\} \qquad \text{and} \qquad [a, b] = \{x \in \mathbb{R} : a \leqslant x \leqslant b\}$$

   are not open sets.
3. The rectangle

$$(a, b) \times (c, d) = \{(x, y) \in \mathbb{R}^2 : a < x < b,\ c < y < d\}$$

   is an open set.

The antithetical notion to an open set is that of a closed set.

Definition 1.5 (Closed set)**.** A subset $C \subseteq \mathbb{R}^n$ is *closed* if its complement $\mathbb{R}^n \backslash C$ is open.

The empty set $\emptyset$ and the space $\mathbb{R}^n$ are the only sets that are both closed and open at the same time. Intuitively, one should think of a closed set as a set that has no "punctures" or "missing endpoints", i.e., it includes all limiting values of points. For instance, the punctured plane $\mathbb{R}^2\backslash\{(0,0)\}$ is not a closed set.

An example of a closed set is the closed ball.

Definition 1.6 (Closed Ball)**.** Let $\mathbf{a} \in \mathbb{R}^n$ and $r > 0$. The set

$$\overline{B(\mathbf{a}, r)} = \{\mathbf{x} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{a}) \leqslant r\}$$

is called the closed ball of radius r centered at $\mathbf{a}$. It is a closed set.

Example 1.2 (Closed Sets in $\mathbb{R}^n$)**.**

1. The closed interval

$$[a, b] = \{x \in \mathbb{R} : a \leqslant x \leqslant b\}$$

   is a closed set, because its complement $\mathbb{R}\backslash[a, b] = (-\infty, a) \cup (b, \infty)$ is an open set.
2. Infinite intervals with closed boundary $[a, \infty)$ and $(-\infty, b]$ are closed sets.
3. Halfopen intervals such as $[a, b)$ or $(a, b]$ are neither closed nor open sets.
4. Any set consisting of only finitely many points is a closed set.

The following two propositions describe how open and closed sets behave under basic set manipulations such as unions, intersections, and set differences.

Proposition 1.1.

- *If $U \subseteq \mathbb{R}^n$ is open and $C \subseteq \mathbb{R}^n$ is closed then $U\backslash C$ is open.*
- *If $C \subseteq \mathbb{R}^n$ is closed and $U \subseteq \mathbb{R}^n$ is open then $C\backslash U$ is closed.*

Proposition 1.2.

- *If $U_1, \ldots, U_k \subseteq \mathbb{R}^n$ are open then $U_1 \cup \ldots \cup U_k$ and $U_1 \cap \ldots \cap U_k$ are open.*
- *If $C_1, \ldots, C_k \subseteq \mathbb{R}^n$ are closed then $C_1 \cup \ldots \cup C_k$ and $C_1 \cap \ldots \cap C_k$ are closed.*

To better grasp the difference between open sets and closed sets, we introduce the concept of interior points, exterior points, and boundary points.

Definition 1.7 (Interior, Exterior, Boundary Points)**.** Let S be a subset of $\mathbb{R}^n$ and $\mathbf{x}$ a point in $\mathbb{R}^n$.

(i) We call $\mathbf{x}$ an *interior point* of S if there exists $r > 0$ such that the ball $B(\mathbf{x}, r)$ is contained in S.

(ii) We call $\mathbf{x}$ an *exterior point* of S if there exists $r > 0$ such that the ball $B(\mathbf{x}, r)$ has empty intersection with S.

(iii) We call $\mathbf{x}$ a *boundary point* of S if it is neither an interior point nor an exterior point for S. Equivalently, $\mathbf{x}$ is a boundary point of S if for every $r > 0$ the ball $B(\mathbf{x}, r)$ has non-empty intersection with S without being entirely contained in S.

Note that every point is either interior, exterior or on the boundary in relationship to a set S.

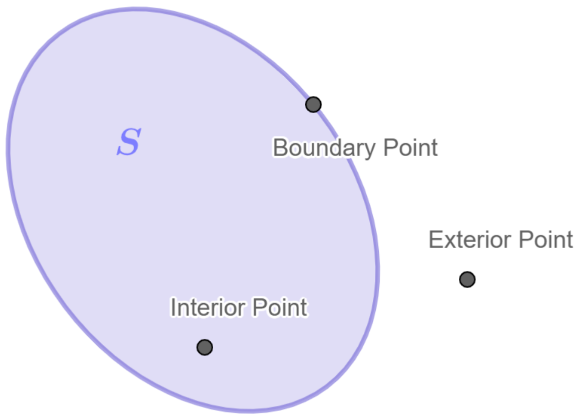


Figure 1.1: Illustration of the difference between interior, exterior and boundary points of a set S.

Definition 1.8 (Interior)**.** The set of all interior points of a set S is called the interior of S and it is denoted by $\mathring{S}$.

Definition 1.9 (Boundary)**.** The set of all boundary points of a set S is called the boundary of S and we use ∂S to denote it.

Definition 1.10 (Closure)**.** The closure of S, denoted by $\overline{S}$, is the set of points $\mathbf{x} \in \mathbb{R}^n$ with the property that for all $r > 0$ one has

$$B(\mathbf{x}, r) \cap S \neq \emptyset.$$

Equivalently, the closure of S is the union of all its interior points and all its boundary points.

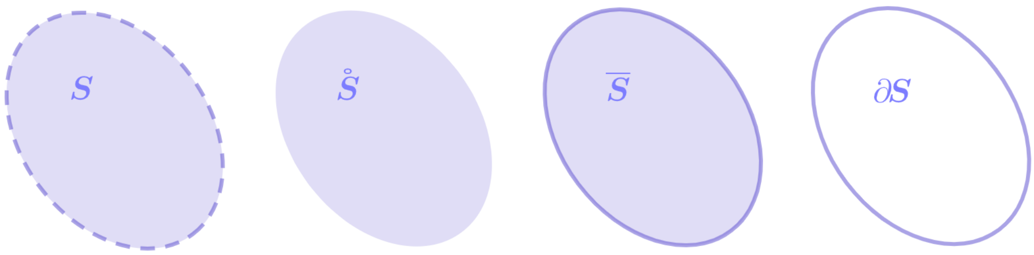


Figure 1.2: The interior, closure and boundary sets of a set S.

Clearly, we have the set inclusions $\mathring{S} \subseteq S \subseteq \overline{S}$. To summarize, the closure of S is S plus its boundary, its interior is S minus its boundary, and the boundary is the closure minus the interior:

$$\mathring{S} = S \backslash \partial S \qquad \overline{S} = S \cup \partial S, \qquad \text{and} \qquad \partial S = \overline{S} \backslash \mathring{S}.$$

Proposition 1.3. *Let $S \subseteq \mathbb{R}^n$. The interior $\mathring{S}$ is the largest open set contained inside of S. The closure $\overline{S}$ is the smallest closed set that has S as a subset.*

Corollary 1.1. *A set is open if and only if it is equal to its interior. On the other hand, a set is closed if and only if it is equal to its closure, which is the same as saying that it contains all its boundary points.*

Example 1.3 (Closure, Interior, Boundary)**.**

1. The sets $(0,1)$, $[0,1]$, $[0,1)$, and $(0,1]$ all have the same closure, interior, and boundary: the closure is $[0,1]$, the interior is $(0,1)$, and the boundary consists of the two points 0 and 1.
2. The sets

$$\{(x,y)\in\mathbb{R}^2 : x^2+y^2<1\} \qquad \text{and} \qquad \{(x,y)\in\mathbb{R}^2 : x^2+y^2\leqslant 1\}$$

both have the same closure, interior, and boundary: the closure is the disc of equation $x^2+y^2\leqslant 1$, the interior is the disc of equation $x^2+y^2<1$, and the boundary is the circle of equation $x^2+y^2=1$.
3. The set

$$U=\{(x,y)\in\mathbb{R}^2 : |y|<x^2\}$$

describes the region between two parabolas touching at the origin, shown in Fig. 1.3. The set is open, so $U=\mathring{U}$. The closure of U is given by

$$\overline{U}=\{(x,y)\in\mathbb{R}^2 : |y|\leqslant x^2\}.$$

In particular, the closure contains the point $(0,0)$.

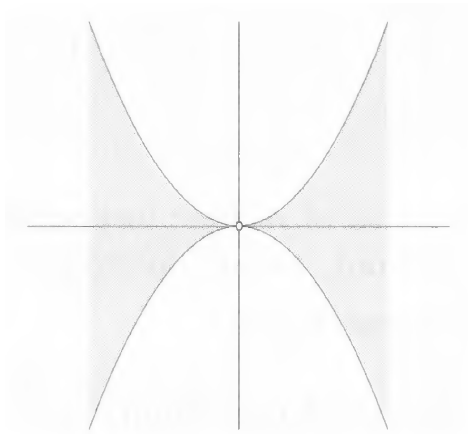


Figure 1.3: The origin belongs to the closure of the shaded region.

4. The unit ball is open in $\mathbb{R}^n$ and is defined by

$$B_1=B(\mathbf{0},1)=\{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}\|_2<1\}$$

Its boundary is the sphere $\partial B_1=\{\mathbf{x}\in\mathbb{R}^n : \|\mathbf{x}\|_2=1\}$.
5. Let $f\colon\mathbb{R}\to\mathbb{R}$ be a continuous function. The set

$$G_f=\{(x,f(x))\in\mathbb{R}^2 : x\in\mathbb{R}\}$$

is known as the graph of f and represents a curve in $\mathbb{R}^2$. We have $\mathring{G}_f = \emptyset$. Therefore $G_f = \partial G_f$. The closed graph theorem says that graph $\mathring{G}_f$ is a closed set in $\mathbb{R}^2$ if f is a continuous function.

6. Let $B = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 < 1\}$ and $I = [0, 5]$. The set S defined by

$$S = B \times I = \left\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1 \text{ and } 0 \leqslant x_3 \leqslant 5\right\}$$

is a cylinder. The set S is neither closed nor open. The boundary of S is given by

$$\partial S = \underbrace{\partial B \times I}_{E_1} \cup \underbrace{B \times \partial I}_{E_2},$$

where

$$\begin{aligned} E_1 &= \left\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 = 1 \text{ and } 0 \leqslant x_3 \leqslant 5\right\}, \\ E_2 &= \left\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1 \text{ and } x_3 \in \{0, 5\}\right\}. \end{aligned}$$

Definition 1.11 (Neighborhood of a point in $\mathbb{R}^n$)**.** Let $\mathbf{x} \in \mathbb{R}^n$ and $U \subseteq \mathbb{R}^n$. If $\mathbf{x}$ is an interior point of U then U is called a *neighborhood of* $\mathbf{x}$.

1.4 Sequences in $\mathbb{R}^n$

Limits of sequences and limits of functions are fundamental notions in calculus, as you already have seen in Analysis 1. Let us extend these principles to higher dimensions. We write $\mathbb{N} = \{1, 2, 3, \ldots\}$ for the set of natural numbers.

Definition 1.12 (Sequences in $\mathbb{R}^n$)**.** A *sequence* of elements of $\mathbb{R}^n$ is a function $k \mapsto \mathbf{x}_k$ that associates to every natural number $k \in \mathbb{N}$ an element $\mathbf{x}_k \in \mathbb{R}^n$. We write $(\mathbf{x}_k)_{k\in\mathbb{N}}$ to denote a sequence in $\mathbb{R}^n$.

Although $(\mathbf{x}_k)_{k\in\mathbb{N}}$ is by definition a sequence of n-tuples, we can also think of it as an n-tuple of sequences by considering each coordinate as an individual sequence,

$$(\mathbf{x}_k)_{k\in\mathbb{N}} = \begin{pmatrix} (x_{1,k})_{k\in\mathbb{N}} \\ \vdots \\ (x_{n,k})_{k\in\mathbb{N}} \end{pmatrix}.$$

Definition 1.13 (Convergent sequence)**.** A sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ of points in $\mathbb{R}^n$ converges to a point $\mathbf{x} \in \mathbb{R}^n$ if for every $\varepsilon > 0$ there exists $N > 1$ such that when $k \geqslant N$, then $d(\mathbf{x}_k, \mathbf{x}) < \varepsilon$. In this case we call $\mathbf{x}$ the *limit* of $(\mathbf{x}_k)_{k\in\mathbb{N}}$ and write

$$\lim_{k\to+\infty} \mathbf{x}_k = \mathbf{x}.$$

Note that not every sequence has a limit, but if a sequence does then this limit is unique. Sequences that possess a limit are called *convergent*, whereas sequences that don't possess one are called *divergent*.

It follows from Definition 1.13 that a sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ converges to $\mathbf{x}$ if and only if the sequence of distances $d(\mathbf{x}_k, \mathbf{x})$ converges to 0, i.e.,

$$\lim_{k\to+\infty} \mathbf{x}_k = \mathbf{x} \quad \Longleftrightarrow \quad \lim_{k\to+\infty} d(\mathbf{x}_k, \mathbf{x}) = 0.$$

Convergence is also observed coordinate wise: A sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ converges to $\mathbf{x}$ if and only if each coordinate of $(\mathbf{x}_k)_{k\in\mathbb{N}}$ converges to the respective coordinate of $\mathbf{x}$. More precisely, if

$$(\mathbf{x}_k)_{k\in\mathbb{N}} = \begin{pmatrix} (x_{1,k})_{k\in\mathbb{N}} \\ \vdots \\ (x_{n,k})_{k\in\mathbb{N}} \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

then

$$\lim_{k\to+\infty} \mathbf{x}_k = \mathbf{x} \quad \Longleftrightarrow \quad \lim_{k\to+\infty} x_{i,k} = x_i \ \text{ for all } i = 1, \ldots, n.$$

Example 1.4 (Convergent and divergent sequences in $\mathbb{R}^n$)**.**

1. The sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ given by

$$\mathbf{x}_k = \begin{pmatrix} e^{-k} \\ \frac{k}{k+1} \\ \frac{1}{\sqrt{k^2-k}-k} \end{pmatrix}$$

   converges as $k \to +\infty$ to the limit

$$\mathbf{x} = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix},$$

   because $\lim_{k\to+\infty} e^{-k} = 0$, $\lim_{k\to+\infty} \frac{k}{k+1} = 1$, and $\lim_{k\to+\infty} \frac{1}{\sqrt{k^2-k}-k} = -2$.

2. The sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ given by

$$\mathbf{x}_k = \begin{pmatrix} 0 \\ \frac{1-(-1)^k}{2} \end{pmatrix}$$

   diverges because it diverges in the second coordinate.

The following properties describe the arithmetic operations of sequences in the n-dimensional Euclidean space and tell us that limits cooperate nicely with the vector space structure of $\mathbb{R}^n$.

Properties of limits of sequences. Let $(\mathbf{x}_k)_{k\in\mathbb{N}}$ and $(\mathbf{y}_k)_{k\in\mathbb{N}}$ be sequences in $\mathbb{R}^n$ and let $(\lambda_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}$.

1. **Addition of sequences:** If $(\mathbf{x}_k)_{k\in\mathbb{N}}$ and $(\mathbf{y}_k)_{k\in\mathbb{N}}$ both converge then so does $(\mathbf{x}_k + \mathbf{y}_k)_{k\in\mathbb{N}}$ and

$$\lim_{k\to+\infty} \mathbf{x}_k + \mathbf{y}_k = \lim_{k\to+\infty} \mathbf{x}_k + \lim_{k\to+\infty} \mathbf{y}_k.$$

2. **Multiplication of sequences:** If $(\mathbf{x}_k)_{k\in\mathbb{N}}$ and $(\lambda_k)_{k\in\mathbb{N}}$ both converge then so

does $(\lambda_k \mathbf{x}_k)_{k\in\mathbb{N}}$ and

$$\lim_{k\to+\infty} \lambda_k \mathbf{x}_k = \left(\lim_{k\to+\infty} \lambda_k\right) \cdot \left(\lim_{k\to+\infty} \mathbf{x}_k\right).$$

3. **Inner product of sequences:** If $(\mathbf{x}_k)_{k\in\mathbb{N}}$ and $(\mathbf{y}_k)_{k\in\mathbb{N}}$ both converge then so does $(\langle \mathbf{x}_k, \mathbf{y}_k\rangle)_{k\in\mathbb{N}}$ and

$$\lim_{k\to+\infty} \langle \mathbf{x}_k, \mathbf{y}_k\rangle = \left\langle \lim_{k\to+\infty} \mathbf{x}_k, \lim_{k\to+\infty} \mathbf{y}_k \right\rangle.$$

Definition 1.14 (Cauchy sequences)**.** A sequence $(\mathrm{x}_k)_{k\in\mathbb{N}}$ is a *Cauchy sequence* if for every $\varepsilon > 0$ there exists $N > 1$ such that $k, l \geqslant N$ implies $d(\mathbf{x}_k, \mathbf{x}_l) < \varepsilon$.

Theorem 1.1. *Every convergent sequence $(\mathrm{x}_k)_{k\in\mathbb{N}}$ is a Cauchy sequence and every Cauchy sequence is convergent.*

Proposition 1.4. *Let $S \subseteq \mathbb{R}^n$ be a non-empty set and suppose $\mathbf{x} \in \partial S$ is a boundary point of S. Then there exists a sequence of elements in $\mathring{S}$, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots \in \mathring{S}$, such that*

$$\lim_{k\to+\infty} \mathbf{x}_k = \mathbf{x}.$$

The following example provides an illustration of the content of Proposition 1.4.

Example 1.5. Consider the open ball of radius 5 centered at the origin in $\mathbb{R}^2$,

$$B(\mathbf{0}, 5) = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 < 5\} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 25\}.$$

The boundary of $B((0,0), 5)$ is the circle of radius 5 centered at the origin, i.e.,

$$\partial B(\mathbf{0}, 5) = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 = 5\} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 25\}.$$

Any point $\mathbf{x} \in \partial B(\mathbf{0}, 5)$ of this circle takes the form

$$\mathbf{x} = \begin{pmatrix} 5\cos\theta \\ 5\sin\theta \end{pmatrix}, \quad \text{for some } \theta \in [0, 2\pi).$$

We can define a sequence

$$\mathbf{x}_k = \begin{pmatrix} \frac{5k}{k+1}\cos\theta \\ \frac{5k}{k+1}\sin\theta \end{pmatrix},$$

and note that $\lim_{k\to+\infty} \mathbf{x}_k = \mathbf{x}$. So we see that $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ is a sequence of points inside the open ball $B(\mathbf{0}, 5)$ converging to the point $\mathbf{x}$ on the border .

Proposition 1.5. *Let $S \subseteq \mathbb{R}^n$ be a non-empty subset of $\mathbb{R}^n$ and let $(\mathbf{x}_k)_{k\in\mathbb{N}}$ be a sequence of elements in S. If $(\mathbf{x}_k)_{k\in\mathbb{N}}$ converges then the limit $\lim_{k\to+\infty} \mathbf{x}_k = \mathbf{x}$ must belong to $\overline{S}$, the closure of S.*

Example 1.6. Consider the "halfopen" rectangle

$$S = [0, 1] \times [0, 1).$$

This is not a closed set, because the point $(\frac{2}{3}, 1)$, for example, is in the boundary ∂S but not in S itself. Moreover, the sequence

$$\begin{pmatrix} \frac{2}{3} \\ \frac{1}{2} \end{pmatrix}, \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}, \begin{pmatrix} \frac{2}{3} \\ \frac{3}{4} \end{pmatrix}, \begin{pmatrix} \frac{2}{3} \\ \frac{4}{5} \end{pmatrix}, \begin{pmatrix} \frac{2}{3} \\ \frac{5}{6} \end{pmatrix}, \dots$$

is a sequence of points in the interior of S that converge to the point $(\frac{2}{3}, 1)$, which is not part of S, but it is part of the closure of S.

Definition 1.15 (Bounded set)**.** A subset $E \subseteq \mathbb{R}^n$ is *bounded* if it is contained in a ball of finite radius centered at the origin:

$$E \subseteq B(\mathbf{0}, R) \quad \text{for some } R < \infty.$$

Note that a closed set need not be bounded. For instance, the interval $[0, \infty)$ is closed, but it is not a bounded.

Definition 1.16 (Compact set)**.** A subset $C \subseteq \mathbb{R}^n$ is *compact* if it is closed and bounded.

Compactness is the basic "finiteness criterion" for subsets of $\mathbb{R}^n$. An important characterization of compact sets in Euclidean spaces is given by the Bolzano-Weierstrass theorem. Before we can state this theorem, we need to recall what is a subsequence.

Definition 1.17 (Subsequence)**.** A *subsequence* of a sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ is any sequence of the form $(\mathbf{x}_{k_i})_{i\in\mathbb{N}}$, where $(k_i)_{i\in\mathbb{N}}$ is a strictly increasing sequence of positive integers.

If a sequence converges then any subsequence of it also converges to the same limit.

Theorem 1.2 (Bolzano-Weierstrass theorem in $\mathbb{R}^n$)**.** *Let $C \subseteq \mathbb{R}^n$ be compact. Any sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ of elements in C possesses a convergent subsequence $(\mathbf{x}_{k_i})_{i\in\mathbb{N}}$ whose limit is in C.*

Definition 1.18 (Bounded sequences in $\mathbb{R}^n$)**.** A sequence $(\mathrm{x}_k)_{k\in\mathbb{N}}$ is *bounded* if there exists a constant $C > 0$ such that $\|\mathbf{x}_k\|_2 \leqslant C$ for any $k \in \mathbb{N}$.

Note that every convergent sequence is a bounded sequence, but the opposite is in general not true. For example, the sequence $x_k = (-1)^k$ is bounded and does not converge. The following is an immediate corollary of the Bolzano-Weierstrass theorem.

Corollary 1.2. *Each bounded sequence $(\mathrm{x}_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ has a convergent subsequence $(\mathbf{x}_{k_i})_{i\in\mathbb{N}}$.*

Chapter 2

Real-valued functions in $\mathbb{R}^n$

Multivariable calculus, also known as *multivariate calculus*, is the extension of calculus in one variable to calculus with functions of several variables. We start by defining real-valued functions in more than one variable.

2.1 Definition

Definition 2.1 (Real-valued function on $E \subseteq \mathbb{R}^n$)**.** Let E be a non-empty subset of $\mathbb{R}^n$. A function $f\colon E \to \mathbb{R}$ that assigns to every element $\mathbf{x} \in E$ a unique real number $y = f(\mathbf{x})$ is called a *real-valued function* on E.

Given a function $f\colon E \to \mathbb{R}$, the *domain* of f is E, denoted $\operatorname{dom}(f)$ or $\operatorname{dom} f$. In theory, the domain should always be a part of the definition of the function rather than a property of it, but in practice it is often the case that the domain is inferred by the description of the function (see Examples 2.1 and 2.3 below).

The *image* (sometimes also called the *range*) of a function f is the set of all the output values that f produces. We denote it by $\operatorname{Im}(f)$ and it is formally defined as

$$\operatorname{Im}(f) = \{f(\mathbf{x}) : \mathbf{x} \in E\} = \{y \in \mathbb{R} : \exists \mathbf{x} \in E \text{ with } f(\mathbf{x}) = y\}.$$

Example 2.1. Let us find and sketch the domain of the function

$$f(x, y) = \frac{\sqrt{x + y + 1}}{(x - 1)}.$$

The expression for f makes sense if the denominator is not 0 and the quantity under the square root sign is nonnegative. So the domain of f is:

$$\operatorname{dom}(f) = \{(x, y) \in \mathbb{R}^2 : x + y + 1 \geqslant 0, x \neq 1\}.$$

The inequality $x + y + 1 > 0$, or $y > -x - 1$, describes the points that lie on or above the line $y = -x - 1$, while $x \neq 1$ means that the points on the line $x = 1$ must be excluded from the domain. See Fig. 2.1 for a sketch of this region.

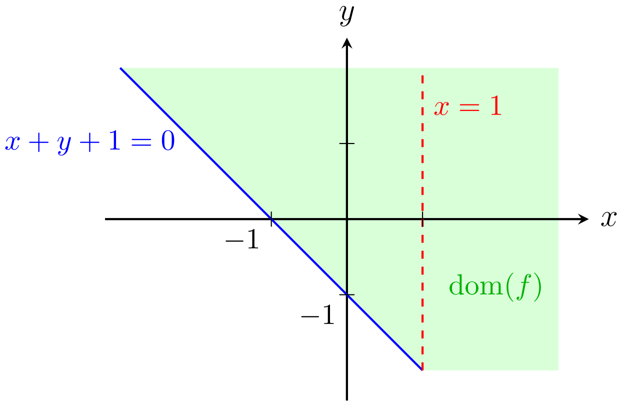


Figure 2.1: The domain of the function $f(x, y) = \frac{\sqrt{x+y+1}}{(x-1)}$
.

The relationship between the domain and the image of a function is described by its *graph*. We use $G(f)$ to denote the graph of a function $f : E \to \mathbb{R}$ and it is given by

$$G(f) = \left\{ \begin{pmatrix} \mathbf{x} \\ f(\mathbf{x}) \end{pmatrix} : \mathbf{x} \in D \right\} \subseteq \mathbb{R}^{n+1}.$$

Note that the graph of f is a subset of $\mathbb{R}^{n+1}$. More precisely, the graph is the hypersurface in $\mathbb{R}^{n+1}$ corresponding to the set of all points $(x_1, \dots, x_n, x_{n+1})^\top \in \mathbb{R}^{n+1}$ that satisfy the equation

$$x_{n+1} = f(x_1, \dots, x_n).$$

Example 2.2. Consider the equation $x + y = z$; as you learned in linear algebra, the solutions to this equation describe a plane in $\mathbb{R}^3$. Now, compare this with the function $f(x, y) = x + y$, a real-valued function in two variables. By definition, the graph of $f(x, y)$ consists of points $(x, y, z) \in \mathbb{R}^3$ where $z = f(x, y)$. For $f(x, y) = x + y$, this gives the equation of the plane $x + y = z$. Thus, the graph of $f(x, y) = x + y$ is exactly the plane in $\mathbb{R}^3$ determined by the equation $x + y = z$.

Example 2.2 connects what you studied in linear algebra, where you worked with linear equations like $x + y = z$, to what you're learning now in multivariable calculus. But there's more! With multivariable functions, you can describe not just planes, but much more complex geometric surfaces, as this next example illustrates.

Example 2.3. Consider the real-valued function $f(x, y) = \sqrt{1 - x^2 - y^2}$, which is a function in 2 variables. The natural domain of this function is $\mathrm{dom}(f) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leqslant 1\}$, which is the closed disc of radius 1 centered at the origin. The image of f is $\mathrm{Im}(f) = [0, 1]$ and the graph $G(f) = \{(x, y, z) \in D \times \mathbb{R}, z = f(x, y)\}$ coincides with the set of solutions to the equations

$$x^2 + y^2 + z^2 = 1 \quad \text{and} \quad z \geqslant 0.$$

In other words, the graph of the function is a *semi-sphere*, see Fig. 2.2 below.

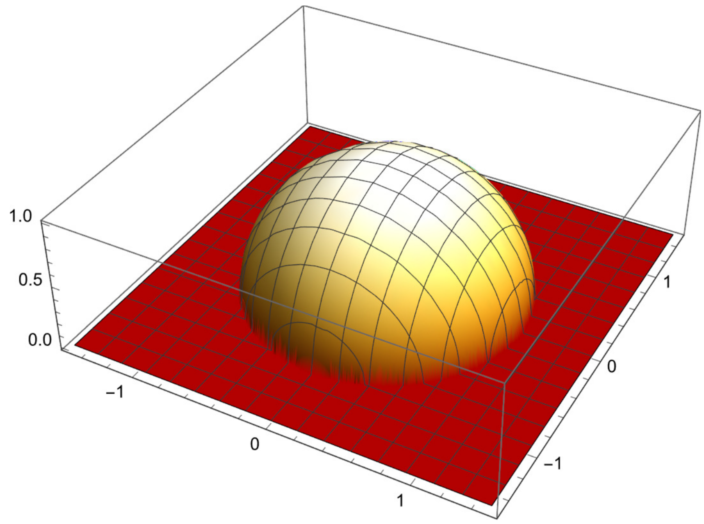


Figure 2.2: Graph of the function $f(x, y) = \sqrt{1 - x^2 - y^2}$.

Example 2.4. In physics, the functions $f\colon \mathbb{R}^n \to \mathbb{R}$ are often called scalar fields. The gravitational potential of a mass or the electric potential of an electric charge are examples of scalar fields:

$$\phi : \mathbb{R}^3 \backslash \{\mathbf{0}\} \to \mathbb{R}, \quad \phi(\mathbf{x}) = \frac{k}{\|\mathbf{x}\|_2}$$

for a real constant k. In mechanics, we often consider systems where the energy is conserved (Hamiltonian systems). For the movement of a particle of mass m in space, subject to the potential $V(\mathbf{x})$, its energy is a real-valued function of its momentum $\mathbf{p} = m\mathbf{v}$ here $\mathbf{v}$ is the velocity and $\mathbf{x}$ the position in space:

$$E : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad E(\mathbf{p}, \mathbf{x}) = \frac{\|\mathbf{p}\|_2^2}{2m} + V(\mathbf{x}).$$

The movement follows the lines at which the energy E is constant. These lines are called “contour lines” and they are special cases of so-called *level sets*, which we define and discuss next.

2.2 Level Sets

Definition 2.2 (Level set)**.** Let $f\colon E \to \mathbb{R}, E \subseteq \mathbb{R}^n (E \neq \emptyset)$. Given a real number $c \in \mathrm{Im}(f)$, we call the set

$$L_c(f) = \{\mathbf{x} \in D : f(\mathbf{x}) = c\} = f^{-1}(\{c\})$$

the *level set* of f at height c. If $c \notin Im(f)$, then $L_c(f) = \emptyset$.

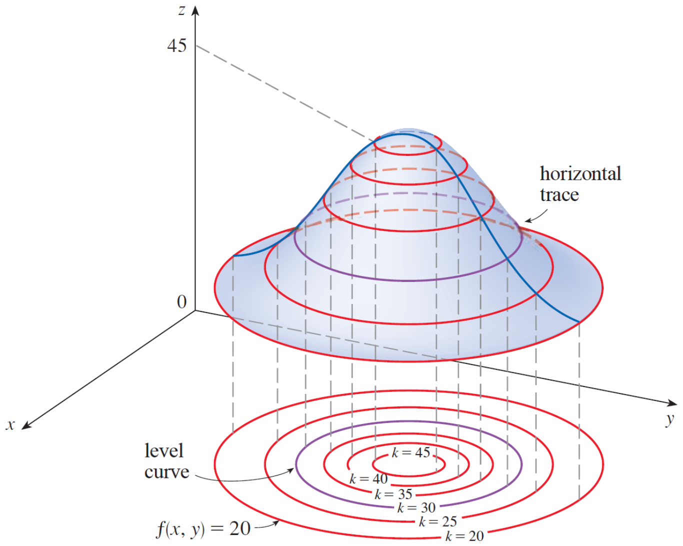


Figure 2.3: The figure displays the graph of a function in 2 variables together with an illustration of its level curves in the xy-plane. One can also think of level curves as the projection of the horizontal traces onto the xy-plane, where a *horizontal trace* is a line formed by intersecting the graph of the function with a plane parallel to the xy-plane.

Level sets of functions in 2 variables $f\colon \mathbb{R}^2 \to \mathbb{R}$ are sometimes also called *level curves* (or *contour lines*). It represents all the points where f has *"height" c*. A collection of contour lines is called a *contour map*. Contour maps are very helpful for visualizing functions, and they are most descriptive if the level curves are drawn for equally spaced heights, see Fig. 2.4.

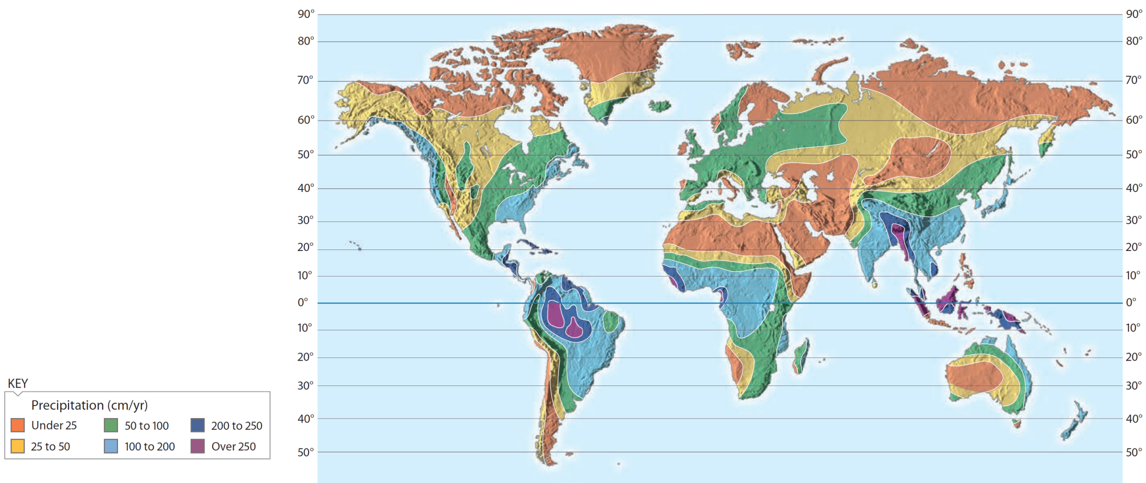


Figure 2.4: Contour map of participation as a function in two variables, the longitude and latitude coordinates on earth.

In summary, we now have learned of two ways of graphically representing a real-valued functions in two variables. The first way is by its graph, which is a hypersurface in $\mathbb{R}^3$, and the second is by a contour map, the projection of its contour lines onto the plane $\mathbb{R}^2$. In Fig. 2.5 below you can see these two methods juxtaposed.

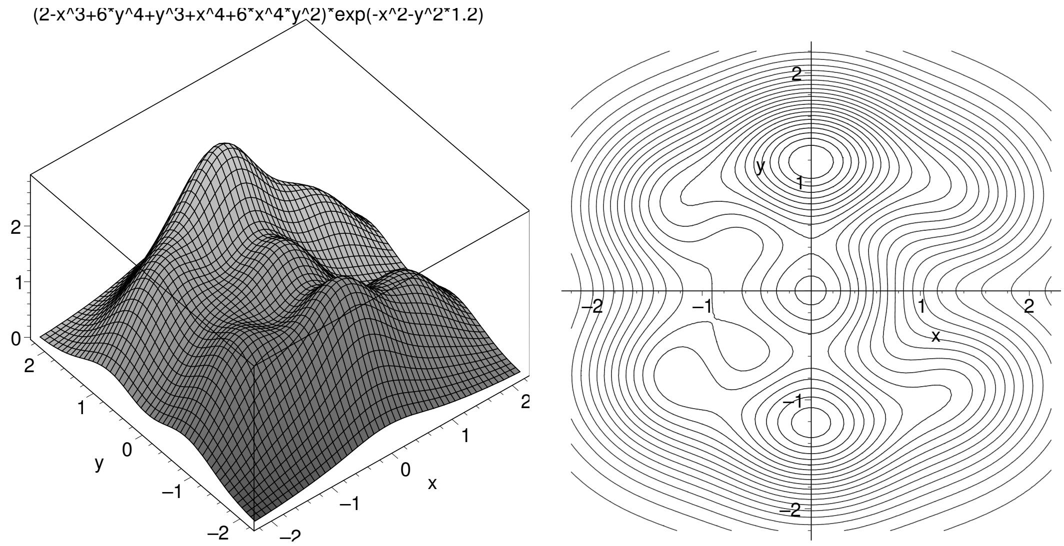


Figure 2.5: Depiction of graph (left) and contour diagram (right) of the same function in 2 variables.

Example 2.5. Let $f(x,y) = \frac{xy-1}{\sqrt{y-x^2}}$, whose domain is $\text{dom}(f) = \{(x,y) \in \mathbb{R}^2 : y > x^2\}$. Notice that $\text{dom}(f)$ is open and unbounded.

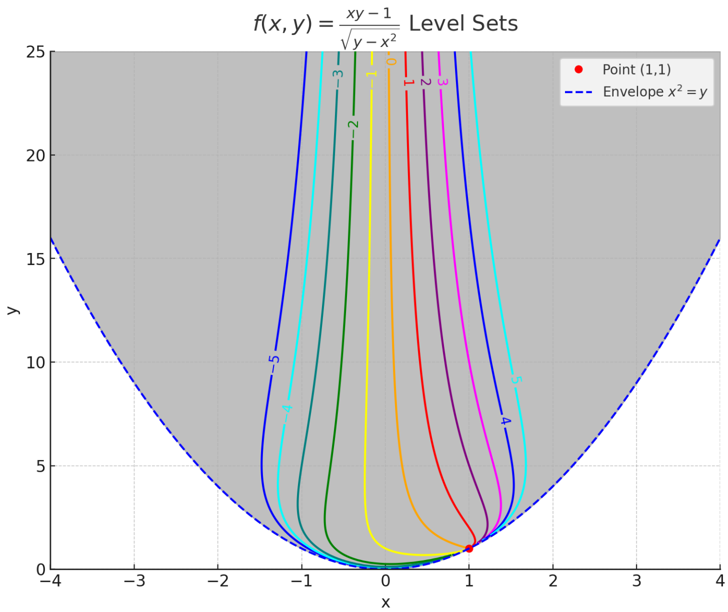


Figure 2.6: The figure displays a series of level curves for the function $f(x,y) = \frac{xy-1}{\sqrt{y-x^2}}$ passing through the point $(1,1)$. As we will explore subsequently, this indicates that the limit of $f(x,y)$ as (x,y) approaches (1,1) is not well-defined.

2.3 Limits of functions

Definition 2.3. Let $f\colon E \to \mathbb{R}$ with $E \subseteq \mathbb{R}^n$. We say that f is *defined in a neighborhood of* $\mathbf{x}_0 \in \mathbb{R}^n$ if $\mathbf{x}_0$ is an interior point of $E \cup \{\mathbf{x}_0\}$. That is, there exists $\delta > 0$ such that $B(\mathbf{x}_0, \delta) \subseteq E \cup \{\mathbf{x}_0\}$.

In the above definition, it is possible that $\mathbf{x}_0 \notin E$. In other words, it is possible for a function to be defined in a neighborhood of $\mathbf{x}_0 \in \mathbb{R}^n$ without being defined at $\mathbf{x}_0$ itself.

Example 2.6. Consider the function $f(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|}$ whose domain equals $\mathrm{dom}(f) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \neq 0\} = \mathbb{R}^n \backslash \{\mathbf{0}\}$. Although this function is not defined at $\mathbf{0}$, it is defined in a neighborhood of $\mathbf{0}$.

We are concerned with points where the function is defined in a neighborhood around the point, because this is necessary to properly define the limit of a function at that point. If the function is not defined in the neighborhood of a point, then it is not always possible to talk about the limit of the function at that point without running into mathematical contradictions.

Definition 2.4 (Limit of a function)**.** Let E be a subset of $\mathbb{R}^n$, $f\colon E \to \mathbb{R}$ a function with domain E and assume f is defined in a neighborhood of the point $\mathbf{x}_0 \in \mathbb{R}^n$. We say that f has a *limit* $l \in \mathbb{R}$ at $\mathbf{x}_0$ and write

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = l,$$

if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $\mathbf{x} \in E$,

$$0 < d(\mathbf{x}, \mathbf{x}_0) \leqslant \delta \implies |f(\mathbf{x}) - l| \leqslant \varepsilon$$

Note that the limit of a function at a point does not always exist. But if it does exists then it is unique, which means that a function has at most one limit at a given point.

Example 2.7. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

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

Let's calculate its limit as (x, y) approaches $(0, 0)$. We will learn several different methods of finding the limit of a function at a point (see, for example, the Squeeze Theorem below), but the most standard method consists of simply verifying Definition 2.4. Given the relation $0 \leqslant \sqrt{x^2+y^2}$, we have

$$\begin{aligned}|f(x,y)| = \frac{|x+y|\,|x^2-xy+y^2|}{x^2+y^2} &\leqslant (|x|+|y|)\frac{x^2+|x||y|+y^2}{x^2+y^2}\\ &\leqslant (|x|+|y|)\frac{x^2+|x||y|+y^2+\frac{1}{2}(|x|-|y|)^2}{x^2+y^2}\\ &= (|x|+|y|)\frac{\frac{3}{2}x^2+\frac{3}{2}y^2}{x^2+y^2}\\ &\leqslant 2\sqrt{x^2+y^2}\frac{\frac{3}{2}x^2+\frac{3}{2}y^2}{x^2+y^2} = 3\sqrt{x^2+y^2} = 3\|(x,y)\|_2.\end{aligned}$$

This shows that as long as $\delta < \frac{\varepsilon}{3}$ we have $d((x,y),(0,0)) < \delta \implies |f(x,y)| \leqslant \varepsilon$. According to Definition 2.4, this means exactly that $\lim_{(x,y)\to(0,0)} f(x,y) = 0$.

Proposition 2.1 (Characterization of limits by sequences)**.** *Let $E \subseteq \mathbb{R}^n, \mathbf{x}_0 \in \mathbb{R}^n$ and assume $f\colon E \to \mathbb{R}$ defined on a neighbourhood of $\mathbf{x_0}$, and $l \in \mathbb{R}^n$. The following statements are equivalent:*

1. $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = l$.
2. $\lim_{k\to\infty} f(\mathbf{x}_k) = l$ *for every sequence* $(\mathbf{x}_k)_{k\in\mathbb{N}}$ *in* $E\backslash\{\mathbf{x}_0\}$ *with* $\lim_{k\to\infty}\mathbf{x}_k = \mathbf{x}_0$.

Properties of limits of functions. Assume $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})$ and $\lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x})$ exist.

1. **Linear combinations:** For constants $\alpha, \beta \in \mathbb{R}$, we have

$$\lim_{\mathbf{x}\to\mathbf{x}_0}(\alpha f(\mathbf{x}) + \beta g(\mathbf{x})) = \alpha\Big(\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})\Big) + \beta\Big(\lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x})\Big)$$

2. **Products:**

$$\lim_{\mathbf{x}\to\mathbf{x}_0}(f(\mathbf{x})\cdot g(\mathbf{x})) = \Big(\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})\Big)\cdot\Big(\lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x})\Big).$$

3. **Quotients:** If $\lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x}) \neq 0$, then

$$\lim_{\mathbf{x}\to\mathbf{x}_0}\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) = \frac{\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x})}{\lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x})}.$$

4. **Compositions:** Let $\mathbf{a} = (a_1,\dots,a_n) \in \mathbb{R}^n$ and $\mathbf{b} = (b_1,\dots,b_n) \in \mathbb{R}^n$ be given. If $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ exists, and $g_i : \mathbb{R} \to \mathbb{R}$ are functions such that $\lim_{x\to b_i} g_i(x) = a_i$ for each i, then

$$\lim_{\mathbf{x}\to\mathbf{b}} f(g_1(x_1), g_2(x_2), \dots, g_n(x_n)) = \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}).$$

Example 2.8. Let us calculate

$$\lim_{(x,y)\to(-3,4)} \frac{1+xy}{1-xy}.$$

Since $\lim_{(x,y)\to(-3,4)} x = -3$ and $\lim_{(x,y)\to(-3,4)} y = 4$, it follows from properties 1 and 2 of limits of functions that

$$\lim_{(x,y)\to(-3,4)} 1+xy \;=\; 1+\Big(\lim_{(x,y)\to(-3,4)} x\Big)\Big(\lim_{(x,y)\to(-3,4)} y\Big) \;=\; 1+(-3)\cdot 4 = -11.$$

Similarly, we obtain $\lim_{(x,y)\to(-3,4)} 1 - xy = 13$. Since the limit of the numerator and denominator exist and the denominator does not converge to 0, it follows from property 3 of limits of functions that

$$\lim_{(x,y)\to(-3,4)} \frac{1+xy}{1-xy} = \frac{\lim_{(x,y)\to(-3,4)} 1+xy}{\lim_{(x,y)\to(-3,4)} 1-xy} = -\frac{11}{13}.$$

2.4 Techniques for finding limits of functions

Example 2.9 (The problem with limits in several variables)**.** Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a function in two variables; we would like to determine the limit

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

A (naïve) idea is to compute the two iterated limits:

$$\lim_{x\to 0}\lim_{y\to 0} f(x,y) \qquad \text{or} \qquad \lim_{y\to 0}\lim_{x\to 0} f(x,y).$$

If these two limits exist and coincide, one might then be led to believe that the limit of the function at $(0,0)$ is equal to 0. However, <u>this is note true</u>! For example, consider the function

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

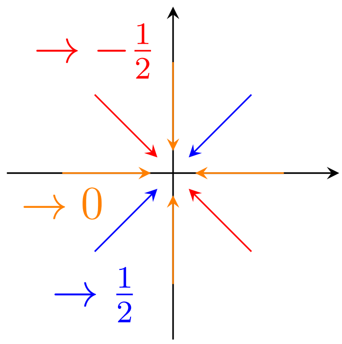
For this particular function, we find that the iterated limits are:

$$\lim_{x\to 0}\lim_{y\to 0} f(x,y) = \lim_{x\to 0}\lim_{y\to 0} \frac{xy}{x^2+y^2} = \lim_{x\to 0} \frac{0}{x^2+0} = 0,$$
$$\lim_{y\to 0}\lim_{x\to 0} f(x,y) = \lim_{y\to 0}\lim_{x\to 0} \frac{xy}{x^2+y^2} = \lim_{y\to 0} \frac{0}{0+y^2} = 0.$$

However, instead having the two variables approach 0 one after the other, we can have them approach zero simultaneously, for example along the diagonal $x = y$. In this case, setting both x and y equal to t and letting t go to zero, we obtain

$$\lim_{t\to 0} f(t,t) = \lim_{t\to 0} \frac{t\cdot t}{t^2+t^2} = \lim_{t\to 0} \frac{1}{2} = \frac{1}{2},$$

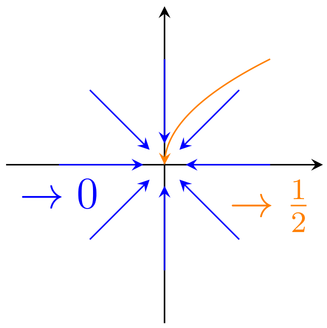
which yields a different result. Since we can approach $(0,0)$ in two different ways and obtain different results, it means that the limit does not exist.



A next idea would be to test all possible directions,

$$\lim_{t\to 0} f(\alpha t, \beta t),$$

with $\alpha, \beta \in \mathbb{R}$ not both zero (thus covering all lines of equation $\beta x - \alpha y = 0$, which are all lines passing through 0). If all the limits along all the lines passing through 0 exist and coincide, can we conclude that the limit exists? The answer is still no! This is because we might obtain a different result when following a trajectory that is not a straight line.



For example, if $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

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

then for any $\alpha, \beta \in \mathbb{R}$, we have

$$\lim_{t\to 0} f(\alpha t, \beta t) = \lim_{t\to 0} \frac{\alpha\beta^2 t^3}{\alpha^2 t^2 + \beta^4 t^4}.$$

If $\alpha = 0$, then $\beta \neq 0$ and we obtain 0. Otherwise,

$$\lim_{t\to 0} f(\alpha t, \beta t) = \lim_{t\to 0} \frac{\alpha\beta^2 t}{\alpha^2 + \beta^4 t^2} = \frac{0}{\alpha + 0} = 0.$$

We obtain 0 in all directions. However,

$$\lim_{t\to 0} f(t^2, t) = \lim_{t\to 0} \frac{t^4}{t^4 + t^4} = \frac{1}{2}.$$

Again, this means that the limit does not exist.

2.4.1 The squeeze theorem

Theorem 2.1 (Squeeze theorem - théorème des gendarmes)**.** *Let $E \subseteq \mathbb{R}^n$, and functions $f, g, h : E \to \mathbb{R}$ be defined on a neighborhood of $\mathbf{x}_0 \in \mathbb{R}^n$. If*

$$\lim_{\mathbf{x}\to\mathbf{x}_0} g(\mathbf{x}) = \lim_{\mathbf{x}\to\mathbf{x}_0} h(\mathbf{x}) = l$$

and there exists $\varepsilon > 0$ such that for all $\mathbf{x} \in E$,

$$0 < d(\mathbf{x}, \mathbf{x}_0) < \varepsilon \implies g(\mathbf{x}) \leqslant f(\mathbf{x}) \leqslant h(\mathbf{x})$$

then

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = l.$$

Example 2.10. Consider $f\colon \mathbb{R}^2\backslash\{(0,0)\} \to \mathbb{R}$ defined by

$$f(x, y) = \frac{x^4 y^3}{x^4 + y^{12}}.$$

Let's discuss the limit

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

We can estimate

$$0 \leqslant f(x, y) = \frac{x^4 y^3}{x^4 + y^{12}} \leqslant \frac{x^4 y^3}{x^4} = y^3.$$

So if we define

$$g(x, y) = 0 \qquad \text{and} \qquad h(x, y) = y^3$$

then $g(x, y) \leqslant f(x, y) \leqslant h(x, y)$. Since $\lim_{(x,y)\to(0,0)} g(x, y) = \lim_{(x,y)\to(0,0)} h(x, y) = 0$, it follows from the Squeeze Theorem that $\lim_{(x,y)\to(0,0)} f(x, y) = 0$.

2.4.2 Using Polar coordinates

Polar coordinates are useful when given a function in two variables involving terms like $x^2 + y^2$, representing the distance from the origin, or when the function behaves similarly along all directions (i.e., has radial symmetry). This simplifies the analysis by converting the problem into one of radial distance and angular symmetry, making it easier to evaluate limits as the distance from the origin approaches zero.

The following version of the squeeze theorem involving polar coordinates allows us to bound a function in terms of its distance from the origin, making it easier to evaluate limits as the distance approaches zero.

Theorem 2.2 (Squeeze theorem in polar coordinates)**.** *Let $E \subseteq \mathbb{R}^2$ and $(x_0, y_0) \in \mathbb{R}^2$. Assume $f\colon E \to \mathbb{R}$ is a function that is defined in the neighborhood of (x_0, y_0) and let*

$l \in \mathbb{R}$. Then,

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = l$$

if and only if there exists $\varepsilon > 0$ and a function $\psi\colon (0,\varepsilon) \to [0,\infty)$ such that

(i) $\lim_{r\to 0^+} \psi(r) = 0$, and

(ii) for all $\theta \in [0,2\pi)$ we have $|f(x_0 + r\cos\theta, y_0 + r\sin\theta) - l| \leqslant \psi(r)$

Example 2.11. Consider $f\colon \mathbb{R}^2\backslash\{(0,0)\} \to \mathbb{R}$ defined by

$$f(x,y) = \frac{x^2 y}{x^2 + y^{\frac{5}{2}}}.$$

Let's discuss the limit

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

We switch to polar coordinates and get

$$\begin{aligned} f(r\cos\theta, r\sin\theta) &= \frac{r^3\cos^2\theta\sin\theta}{r^2\cos^2\theta + r^{\frac{5}{2}}\sin^{\frac{5}{2}}\theta} \\ &= \frac{r\cos^2\theta\sin\theta}{\cos^2\theta + r^{\frac{1}{2}}\sin^{\frac{5}{2}}\theta}. \end{aligned}$$

Thus,

$$|f(r\cos\theta, r\sin\theta)| = \frac{r\cos^2\theta|\sin\theta|}{\cos^2\theta + r^{\frac{1}{2}}\sin^{\frac{5}{2}}\theta} \leqslant \frac{r\cos^2\theta|\sin\theta|}{\cos^2\theta} = r|\sin\theta| \leqslant r.$$

Taking $l = 0$ and $\psi(r) = r$, we see that the hypothesis of the squeeze theorem in polar coordinates is satisfied, and conclude that

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

2.4.3 Using Taylor's theorem

Taylor's theorem (which you have learned in Analysis I) can be useful to find limits because it approximates a function near a point by a polynomial, simplifying the analysis before applying the squeeze theorem. For convenience, let us quickly recall the statement of Taylor's theorem.

Theorem 2.3 (Taylor's theorem – single variable case)**.** *Let $k \in \mathbb{N}$. Suppose $I \subseteq \mathbb{R}$ is an open interval and $f\colon I \to \mathbb{R}$ is a function of class $C^k(I)$. Then for any $a \in I$ we* *have*

$$f(x) = \underbrace{\overbrace{\sum_{j=1}^{k} \frac{f^{(j)}(a)}{j!}(x-a)^k}^{k^{\text{th}}\text{-order approximation}} + \overbrace{r_k(x)}^{\text{remainder}}}_{k^{\text{th}}\text{-order expansion}}$$

where $r_k(x)$ is an "error" term satisfying $\lim_{x\to a} \frac{r_k(x)}{|x-a|^k} = 0$.

Example 2.12. Calculate the following limits if they exist:

(a) $\lim_{(x,y)\to(0,0)} \frac{x^2+\ln(1+y^2)}{\sqrt{x^2+y^2}}$

(b) $\lim_{(x,y)\to(0,0)} \frac{1-e^{(x^3)}}{x^2+y^2}$

(a) The first-order expansion of $\ln(1+x)$ around $a=0$ is

$$\ln(1+x) = x + r_1(x)$$

where $\lim_{x\to 0} \frac{r_1(x)}{x} = 0$. We obtain

$$\lim_{(x,y)\to(0,0)} \frac{x^2+\ln(1+y^2)}{\sqrt{x^2+y^2}} = \lim_{(x,y)\to(0,0)} \frac{x^2+y^2+r_1(y^2)}{\sqrt{x^2+y^2}}$$

$$= \lim_{(x,y)\to(0,0)} \frac{x^2+y^2}{\sqrt{x^2+y^2}} + \lim_{(x,y)\to(0,0)} \frac{r_1(y^2)}{\sqrt{x^2+y^2}} = 0+0=0.$$

The second limit is zero because, for $(x,y)\neq(0,0)$,

$$-\frac{|r_1(y^2)|}{|y|} \leqslant \frac{r_1(y^2)}{\sqrt{x^2+y^2}} \leqslant \frac{|r_1(y^2)|}{|y|}$$

with

$$\lim_{(x,y)\to(0,0)} \frac{|r_1(y^2)|}{|y|} = \lim_{(x,y)\to(0,0)} |y| \cdot \lim_{(x,y)\to(0,0)} \frac{|r_1(y^2)|}{|y^2|} = 0\cdot 0 = 0.$$

By the squeeze theorem, it follows that

$$\lim_{(x,y)\to(0,0)} \frac{r_1(y^2)}{\sqrt{x^2+y^2}} = 0.$$

(b) The first-order expansion of e^x around $a=0$ is

$$e^x = 1 + x + r_1(x)$$

where $\lim_{x\to 0} \frac{r_1(x)}{x} = 0$. We obtain

$$\lim_{(x,y)\to(0,0)} \frac{1-e^{x^3}}{x^2+y^2} = \lim_{(x,y)\to(0,0)} \frac{1-1-x^3-r_1(x^3)}{x^2+y^2} = \lim_{(x,y)\to(0,0)} \frac{-x^3-r_1(x^3)}{x^2+y^2}.$$

Now, for $(x, y) \neq (0,0)$,

$$-\frac{|x^3| + |r_1(x^3)|}{|x^2|} \leqslant \frac{-x^3 - r_1(x^3)}{x^2+y^2} \leqslant \frac{|x^3| + |r_1(x^3)|}{|x^2|}$$

with

$$\lim_{(x,y)\to(0,0)} \frac{|x^3| + |r_1(x^3)|}{|x^2|} = \left(\lim_{(x,y)\to(0,0)} \frac{|x^3|}{|x^2|}\right) + \left(\lim_{(x,y)\to(0,0)} \frac{|r_1(x^3)|}{|x^2|}\right) = 0 + 0 = 0.$$

The squeeze theorem therefore ensures that

$$\lim_{(x,y)\to(0,0)} \frac{-x^3 - r_1(x^3)}{x^2+y^2} = 0.$$

2.4.4 Using change of variables

The following proposition enables us to convert limits in two variables into limits in a single variable.

Proposition 2.2 (Composition with Functions of a Single Variable)**.** *Let $E \subseteq \mathbb{R}^2$ and let $g\colon E \to \mathbb{R}$ be defined in a neighborhood of $(x_0, y_0) \in \mathbb{R}^2$. Let $I \subseteq \mathbb{R}$ be such that $I \subseteq g(E)$ and let $\varphi\colon I \to \mathbb{R}$ be defined in a neighborhood of $l \in \mathbb{R}$. Finally, let $f\colon E \to \mathbb{R}$ be defined by $f(x,y) = \varphi(g(x,y))$. If*

$$\lim_{(x,y)\to(x_0,y_0)} g(x,y) = l \qquad \text{and} \qquad \lim_{t\to l} \varphi(t) \text{ exists,}$$

then

$$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \lim_{t\to l} \varphi(t).$$

Example 2.13. Let $f\colon \mathbb{R}^2 \backslash \{(0,0)\} \to \mathbb{R}$ be defined by

$$f(x,y) = \frac{\tan(3x^2+y^2)}{3x^2+y^2}.$$

We analyze the limit

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

If we define $g(x,y) = 3x^2 + y^2$, then by properties of limits we have

$$\lim_{(x,y)\to(0,0)} g(x,y) = 3\Big(\lim_{(x,y)\to(0,0)} x\Big)^2 + \Big(\lim_{(x,y)\to(0,0)} y\Big)^2 = 3\cdot 0^2 + 0^2 = 0.$$

Define $\varphi\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$ by

$$\varphi(t) = \frac{\tan(t)}{t}.$$

Then we have $f(x,y) = \varphi(g(x,y))$. Hence, in light of Proposition 2.2, we have

$$\lim_{(x,y)\to(0,0)} \frac{\tan(3x^2+y^2)}{3x^2+y^2} = \lim_{t\to 0} \frac{\tan(t)}{t}.$$

Now,

$$\lim_{t\to 0} \frac{\tan(t)}{t} \overset{\text{L'Hôpital's Rule}}{=} \lim_{t\to 0} \frac{\frac{1}{\cos^2(t)}}{1} = 1.$$

Thus,

$$\lim_{(x,y)\to(0,0)} f(x,y) = 1.$$

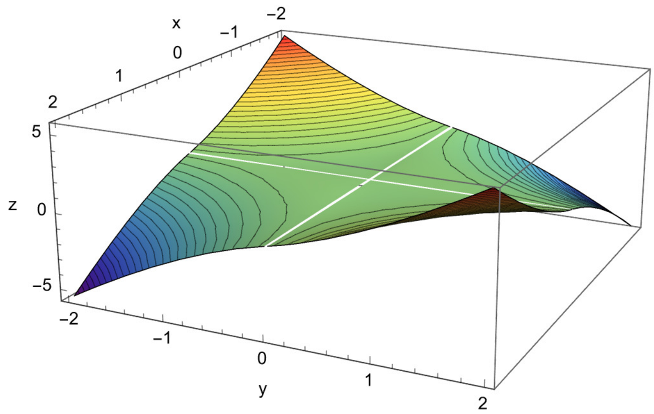


Figure 2.7: Graph of the function $f(x,y) = xy\ln(|x|+|y|)$.

Example 2.14. Let us demonstrate that the limit of the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

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

is zero as (x,y) approaches $(0,0)$ (see Fig. 2.7). Note that for every point (x,y) with $0 < \sqrt{x^2+y^2} < 1$ we have $|xy| \leqslant |x|+|y|$. This implies that for any such (x,y) we have the estimate

$$0 \leqslant |f(x,y)| = |xy\ln(|x|+|y|)| \leqslant (|x|+|y|)|\ln(|x|+|y|)|.$$

So if we define

$$g(x,y) = -(|x|+|y|)|\ln(|x|+|y|)| \qquad \text{and} \qquad h(x,y) = (|x|+|y|)|\ln(|x|+|y|)|$$

then we see that

$$0 < \sqrt{x^2+y^2} < 1 \implies g(x,y) \leqslant f(x,y) \leqslant h(x,y).$$

Substituting t for $|x|+|y|$, it follows from Proposition 2.2 that:

$$\lim_{(x,y)\to(0,0)} \pm(|x|+|y|)|\ln(|x|+|y|)| = \lim_{t\to 0+} t\ln t = 0,$$

where we used the fact $\lim_{t\to 0+} t\ln t = 0$, which can be verified using L'Hôpital's Rule. In other words $\lim_{(x,y)\to(0,0)} g(x,y) = \lim_{(x,y)\to(0,0)} h(x,y) = 0$. Invoking the Squeeze Theorem, we conclude that $\lim_{(x,y)\to(0,0)} f(x,y) = 0$.

2.4.5 Testing along polynomial paths

Testing paths of the form (t^α, t^β) is useful for evaluating limits of functions in two variables because these paths allow us to explore how the function behaves along different directions approaching the origin. By adjusting the exponents α and β, we can test a variety of trajectories that the function might take, revealing whether the limit depends on the direction of approach.

Example 2.15. Let $f : \mathbb{R}^2\backslash\{(0,0)\} \to \mathbb{R}$ be defined by

$$f(x,y) = \frac{x^3y^3}{x^4+y^{12}}.$$

Our goal is to determine the limit

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

First, let us test all linear paths by considering

$$\lim_{t\to 0} f(\alpha t, \beta t),$$

with $\alpha, \beta \in \mathbb{R}$ not both zero. In this case, we get

$$\lim_{t\to 0} f(\alpha t, \beta t) = \lim_{t\to 0} \frac{\alpha^3\beta^3t^6}{\alpha^4t^4+\beta^{12}t^{12}} = \lim_{t\to 0} \frac{\alpha^3\beta^3t^2}{\alpha^4+\beta^{12}t^8} = 0.$$

We see that all linear paths yield the same limit. Therefore, to demonstrate that the limit does not exist, we must consider non-linear paths.

When dealing with a denominator containing different powers of x and y, a good approach is to examine paths of the form (t^α, t^β) for various values of $\alpha, \beta \in (0,\infty)$. This gives

$$\lim_{t\to 0} f(t^\alpha, t^\beta) = \lim_{t\to 0} \frac{t^{3\alpha+3\beta}}{t^{4\alpha}+t^{12\beta}}.$$

First, we can take $\alpha = \beta = 1$. In this case we have

$$\lim_{t\to 0} f(t,t) = \lim_{t\to 0} \frac{t^6}{t^4+t^{12}} = \lim_{t\to 0} \frac{t^2}{1+t^8} = 0.$$

Next, we choose α and β so that the powers appearing in the denominator match. For us, this means we want to find α and β such that

$$4\alpha = 12\beta.$$

For example, this is achieved by taking $\alpha = 3$ and $\beta = 1$. Then,

$$\lim_{t\to 0} f(t^3, t) = \lim_{t\to 0} \frac{t^{12}}{t^{12} + t^{12}} = \frac{1}{2}.$$

Since $\alpha = \beta = 1$ and $\alpha = 3, \beta = 1$ yield different results, we conclude that the limit does not exist.

2.5 Continuity at a Point

The purpose of this section is to introduce and discuss continuous functions in several variables.

Definition 2.5 (Continuous function at a point)**.** Let $E \subseteq \mathbb{R}^n$ and let $\mathbf{x}_0$ be an interior point of E. A function $f\colon E \to \mathbb{R}$ is said to be continuous at $\mathbf{x}_0$ if

$$\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0).$$

Definition 2.6 (1^{st} equivalent definition)**.** Let $\mathbf{x}_0$ be an interior point of E. A function $f\colon E \to \mathbb{R}$ is continuous at $\mathbf{x}_0$ if and only if, for every real number $\varepsilon > 0$, there exists a real number $\delta > 0$ such that for all $\mathbf{x} \in E$,

$$d(\mathbf{x}, \mathbf{x}_0) \leqslant \delta \implies |f(\mathbf{x}) - f(\mathbf{x}_0)| \leqslant \varepsilon.$$

Definition 2.7 (2^{nd} equivalent definition)**.** Let $\mathbf{x}_0$ be an interior point of E. A function $f\colon E \to \mathbb{R}$ is continuous at $\mathbf{x}_0$ if and only if, for every sequence $(\mathbf{a}_k)_{k\in\mathbb{N}}$ of elements of E we have

$$\lim_{k\to+\infty} \mathbf{a}_k = \mathbf{x}_0 \implies \lim_{k\to+\infty} f(\mathbf{a}_k) = f(\mathbf{x}_0).$$

Remark 2.1. It is very tempting to believe that if a function is continuous in every coordinate then the function is continuous. However, this is NOT TRUE! As a counterexample, consider the function

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

Let $f_1, f_2\colon \mathbb{R} \to \mathbb{R}$ denote the two functions obtained by restricting $f(x, y)$ to the first and second coordinate at the point $(0, 0)$, that is, $f_1(x) = f(x, 0)$ and $f_2(y) = f(0, y)$. Then $f_1(x)$ and $f_2(y)$ both are continuous at $x = 0$ and $y = 0$ respectively. Nonetheless, we have already seen in Example 2.9 that the limit of $f(x, y)$ as (x, y) approaches $(0, 0)$ does not exist, which means that the function $f(x, y)$ (as a function in two variables) is not continuous at the point $(0, 0)$.

Properties of continuity. Let f and g be two functions from $E \subseteq \mathbb{R}^n$ to $\mathbb{R}$ that are continuous at a point $\mathbf{x}_0 \in \mathbb{R}^n$. Then:

1. **Linear combinations:** For any $\alpha, \beta \in \mathbb{R}$, the function $\alpha f + \beta g$ is continuous at $\mathbf{x}_0$;
2. **Products:** The product function fg is continuous at $\mathbf{x}_0$;
3. **Quotients:** If $g(\mathbf{x}_0) \neq 0$ and $g(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in E$ then the quotient $\frac{f}{g}$ is continuous at $\mathbf{x}_0$;
4. **Compositions:** Let A be a subset of $\mathbb{R}^n$ and let

$$g_1, \dots, g_p : A \to \mathbb{R}$$

be functions continuous at the point $\mathbf{a} = (a_1, \dots, a_n)$. On the other hand, let B be a subset of $\mathbb{R}^p$ containing

$$\{(g_1(\mathbf{y}), \dots, g_p(\mathbf{y})) : \mathbf{y} \in A\}$$

and $f \colon B \to \mathbb{R}$ a function continuous at the point $\mathbf{b} = (g_1(\mathbf{a}), \dots, g_p(\mathbf{a}))$. Then the function $F \colon A \to \mathbb{R}$ defined by

$$F(y_1, \dots, y_n) = f(g_1(y_1, \dots, y_n), \dots, g_p(y_1, \dots, y_n))$$

is continuous at the point $\mathbf{a} = (a_1, \dots, a_n)$.

Example 2.16. Let us demonstrate the usefulness of the properties of continuity by showing that the function $F \colon \mathbb{R}^2 \to \mathbb{R}$ given by $F(x, y) = -\sin(x)y$ is continuous at the point $(0, 0)$. To do this, consider the three auxiliary functions $f \colon \mathbb{R}^2 \to \mathbb{R}$ and $g_1, g_2 \colon \mathbb{R} \to \mathbb{R}$ defined respectively by

$$f(x, y) = xy, \qquad g_1(x, y) = -\sin(x), \qquad \text{and} \qquad g_2(x, y) = y.$$

Since both $g_1(x, y)$ and $g_2(x, y)$ are continuous at $(0, 0)$ and $f(x, y)$ is continuous at $(g_1(0, 0), g_2(0, 0)) = (0, 0)$, we can conclude that $F(x, y) = f(g_1(x, y), g_2(x, y))$ is continuous at the point $(0, 0)$ (See Fig. 2.8).

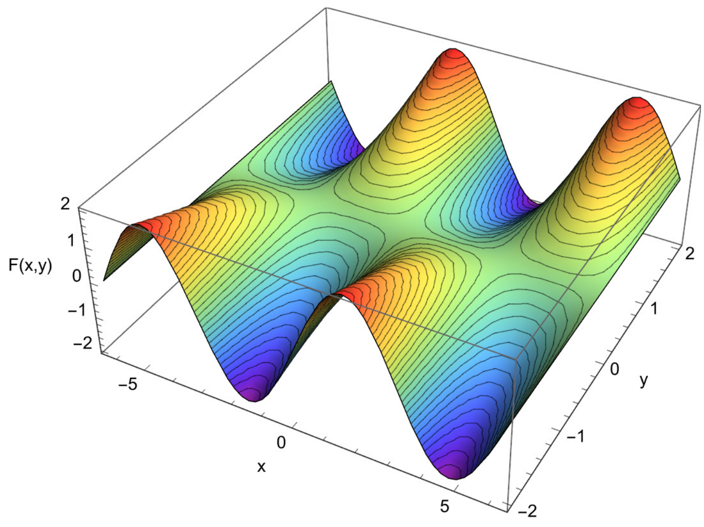


Figure 2.8: Graph of the function $F(x, y) = -\sin(x)y$.

2.6 Continuity in a Region

Definition 2.8 (Continuous function in a Region)**.** Let E be a non-empty subset of $\mathbb{R}^n$. A function $f\colon E \to \mathbb{R}$ is continuous on E if for every $\mathbf{x}_0 \in E$ and every real number $\varepsilon > 0$ there exists a real number $\delta > 0$ such that for all $\mathbf{x} \in E$,

$$d(\mathbf{x}, \mathbf{x}_0) \leqslant \delta \implies |f(\mathbf{x}) - f(\mathbf{x}_0)| \leqslant \varepsilon.$$

Definition 2.9 (Equivalent definition)**.** Let E be a non-empty subset of $\mathbb{R}^n$. A function $f\colon E \to \mathbb{R}$ is continuous on E if for every sequence $(\mathbf{a}_k)_{k\in\mathbb{N}}$ of elements of E we have

$$\lim_{k\to+\infty} \mathbf{a}_k = \mathbf{x}_0 \implies \lim_{k\to+\infty} f(\mathbf{a}_k) = f(\mathbf{x}_0).$$

Remark 2.2. If E is an open set then $f\colon E \to \mathbb{R}$ is continuous on E if and only if it is continuous at every point in E.

Example 2.17. Let us demonstrate that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

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

is continuous on $\mathbb{R}^2$ (see Fig. 2.9). Define the function $h : \mathbb{R} \to \mathbb{R}$ by

$$h(s) = \begin{cases} \frac{\sin(s)}{s} & \text{if } s \neq 0 \\ 1 & \text{if } s = 0 \end{cases}$$

It is continuous for all $s \neq 0$ and, as $\lim_{s\to 0} h(s) = 1 = h(0)$, it is also continuous at 0. This is useful because we have $f(x,y) = h(xy)y$ for all $(x,y) \in \mathbb{R}^2$. Since the functions

$$a(x,y) = xy \qquad \text{and} \qquad b(x,y) = y$$

are continuous at every point in $\mathbb{R}^2$ and $f(x,y) = h(xy)y = a(h(a(x,y)), b(x,y))$ for all $(x,y) \in \mathbb{R}^2$, it follows from the properties of continuity that f is continuous at every point in $\mathbb{R}^2$.

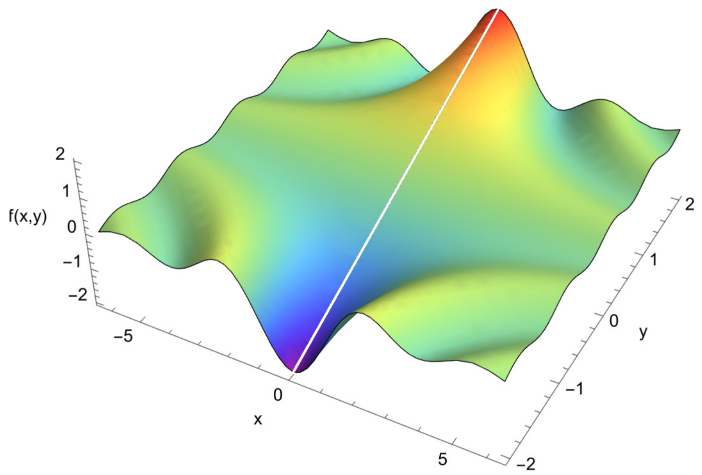


Figure 2.9: Graph of $f(x, y) = \frac{\sin(xy)}{x}$ for $x \neq 0$.

2.7 Extreme Value Theorem and Intermediate Value Theorem

Definition 2.10 (Maximum and minimum)**.** Let $E \subseteq \mathbb{R}^n$ be non-empty and f a function from E to $\mathbb{R}$. A real number M satisfying

- $f(\mathbf{x}) \leqslant M$ for every element $\mathbf{x}$ in E, and
- $M \in \mathrm{Im}(f)$,

is called the *maximum* of the function f on E and is denoted by $\max_{\mathbf{x}\in E} f(\mathbf{x})$. If $\mathbf{x}_0 \in E$ is such that $f(\mathbf{x}_0) = M$ then we say that the function f reaches its maximum at the point $\mathbf{x}_0$. Similarly, a real number m satisfying

- $f(\mathbf{x}) \geqslant m$ for every element $\mathbf{x}$ in E, and
- $m \in \mathrm{Im}(f)$,

is called the *minimum* of the function f on E and is denoted by $\min_{\mathbf{x}\in E} f(\mathbf{x})$. If $\mathbf{x}_0 \in E$ is such that $f(\mathbf{x}_0) = M$ then we say that the function f reaches its minimum at the point $\mathbf{x}_0$.

Proposition 2.3 (Extreme value theorem)**.** *Let E be a compact subset of $\mathbb{R}^n$ and $f \colon E \to \mathbb{R}$ a continuous function. Then f has a minimum $\min_{\mathbf{x}\in E} f(\mathbf{x})$ and a maximum $\max_{\mathbf{x}\in E} f(\mathbf{x})$ on E.*

Chapter 3

Partial derivatives and differentiability

3.1 Partial Derivatives

Recall that given a differentiable function in a single variable $f\colon \mathbb{R} \to \mathbb{R}$, the derivative of f at the point $a \in \mathbb{R}$ is defined as

$$f'(a) = \frac{df}{dx}(a) = \lim_{t\to 0} \frac{f(a+t) - f(a)}{t} = \lim_{x\to a} \frac{f(x) - f(a)}{x - a}.$$

We are already familiar with several different ways of thinking about the derivative of a function:

- The derivative of a function f quantifies the rate of change of the function's output value with respect to its input value. For example, if the derivative of f at a point a is a 'large' positive number then a positive change close to a will result in a 'proportionately large' positive change in the output value. Conversely, if the derivative of f at a point a is a 'small' negative number then a positive change close to a will result in a 'proportionately small' negative change in the output value.
- The derivative $f'(a)$ of a function f at a point a equals the slope of the tangent line to the graph of the function at that point. Moreover, the tangent line is the best linear approximation of the function near that input value.

The goal of this chapter is to extend derivatives to functions in several variables. While functions in one variable have only one derivative, functions in several variables have multiple derivatives, one for each variable. These are called the partial derivatives.

Let

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; \ldots, \; \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

denote the vectors of the canonical basis of $\mathbb{R}^n$. Note that for any element $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ we have $\mathbf{x} = \sum_{k=1}^{n} x_k \mathbf{e}_k$, where $x_k = \langle \mathbf{x}, \mathbf{e}_k \rangle$, for $k = 1, \dots, n$.

Definition 3.1 (Partial derivatives)**.** Suppose $E \subseteq \mathbb{R}^n$ is a set and $\mathbf{a} = (a_1, \dots, a_n)$ is an interior point of E. Let $f \colon E \to \mathbb{R}$ be a real-valued function in the variables $(x_1, \dots, x_n)$. The *partial derivative* of f at the point $\mathbf{a}$ with respect to the variable x_k (the k-th variable) is defined as

$$\frac{\partial f}{\partial x_k}(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{e}_k) - f(\mathbf{a})}{t}$$

whenever this limit exists. If this limit does not exist then we say that the partial derivative of f at $\mathbf{a}$ with respect to x_k does not exist.

Intuitively, the partial derivative $\frac{\partial f}{\partial x_k}$ is the derivative of $f(x_1, \dots, x_n)$ with respect to the variable x_k while all the other variables remain constant. We also use the notation

$$D_k f(\mathbf{a}) = \frac{\partial f}{\partial x_k}(\mathbf{a});$$

or if the real variables of f are explicitly given, say $f(x, y, z)$, then we write

$$\begin{aligned} D_x f(x, y, z) &= \frac{\partial f}{\partial x}(x, y, z) \\ D_y f(x, y, z) &= \frac{\partial f}{\partial y}(x, y, z) \\ D_z f(x, y, z) &= \frac{\partial f}{\partial z}(x, y, z). \end{aligned}$$

Remark 3.1. The partial derivative $\frac{\partial f}{\partial x_k}(\mathbf{a})$ exists if and only if the function $g_k(t) = f(\mathbf{a} + t\mathbf{e}_k)$ is differentiable at $t = 0$, because

$$\frac{\partial f}{\partial x_k}(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{e}_k) - f(\mathbf{a})}{t} = \lim_{t \to 0} \frac{g_k(t) - g_k(0)}{t} = g_k'(0). \tag{3.1}$$

This means that $\frac{\partial f}{\partial x_k}(\mathbf{a})$ corresponds to the slope of the tangent line pointing in the direction of the canonical vector $\mathbf{e}_k$. In the case of two variables, Fig. 3.1 below provides an illustration of the partial derivatives as the slope of tangent lines in the x-direction and in the y-direction.

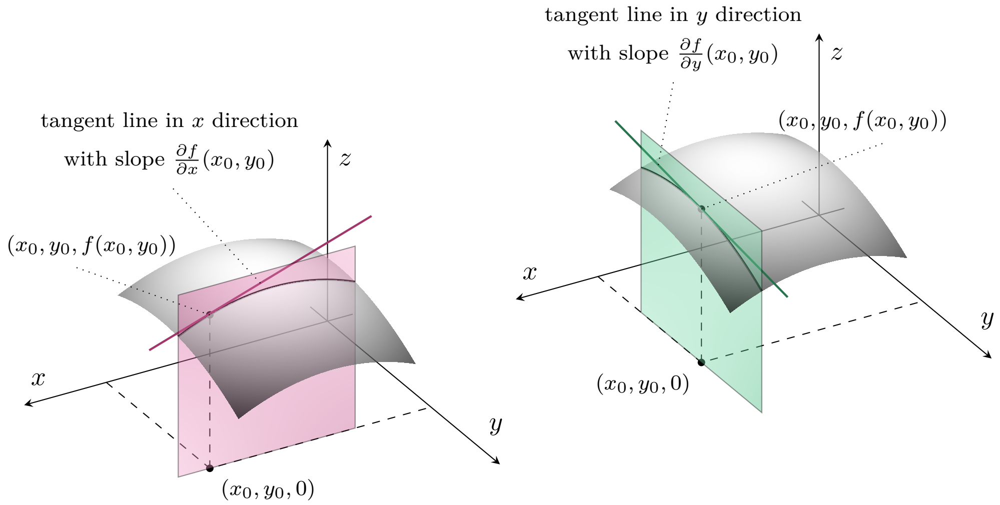


Figure 3.1: The gray surface is the graph of the function $f(x, y)$ and contains the point $(x_0, y_0, f(x_0, y_0))$. In the left figure, the plane $y = y_0$ (pink plane) intersects the graph of $f(x, y)$ in a curve. The tangent line to this curve at the point $(x_0, y_0, f(x_0, y_0))$ (pink line) has slope equal to the partial derivative of $f(x, y)$ with respect to the variable x at the point (x_0, y_0). The right figure depicts the tangent line (green line) to the curve that is the intersection of the graph of $f(x, y)$ with the plane $x = x_0$ (green plane) at the point $(x_0, y_0, f(x_0, y_0))$, whose slope is the partial derivative of $f(x, y)$ with respect to the variable y at the point (x_0, y_0).

Example 3.1. Consider a pot filled with water being heated on top of a stove (see Fig. 3.2). Let us think of the pot as a cylinder in $\mathbb{R}^3$ given by

$$D = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 < 1,\ 0 < z < 1\}.$$

Suppose at time t the temperature of the water at the position (x, y, z) is given by the equation

$$T(x, y, z, t) = \left(100 - \frac{80}{1+t}\right) \cdot \left(1 - \frac{z}{2}\right) \cdot e^{-x^2-y^2}.$$

Then T is a function in 4 variables (3 space variables and 1 time variable) with domain $\operatorname{dom}(T) = D \times [0, \infty)$. We can calculate its partial derivatives as

$$\begin{aligned}
T_x(x, y, z, t) &= \frac{\partial T}{\partial x}(x, y, z, t) = \left(100 - \frac{80}{1+t}\right) \cdot \left(1 - \frac{z}{2}\right) \cdot (-2x) \cdot e^{-x^2-y^2}, \\
T_y(x, y, z, t) &= \frac{\partial T}{\partial y}(x, y, z, t) = \left(100 - \frac{80}{1+t}\right) \cdot \left(1 - \frac{z}{2}\right) \cdot (-2y) \cdot e^{-x^2-y^2}, \\
T_z(x, y, z, t) &= \frac{\partial T}{\partial z}(x, y, z, t) = \left(100 - \frac{80}{1+t}\right) \cdot \left(-\frac{1}{2}\right) \cdot e^{-x^2-y^2}, \\
T_t(x, y, z, t) &= \frac{\partial T}{\partial t}(x, y, z, t) = \frac{80}{(1+t)^2} \cdot \left(1 - \frac{z}{2}\right) \cdot e^{-x^2-y^2}.
\end{aligned}$$

What do these partial derivatives mean? For example, $T_t(x, y, z, t)$ describes the rate of change in temperature at a stationary point (x, y, z) as time t changes. Since T_t is always positive, we see that in every point (x, y, z) the temperature is increasing as the time t increases. Conversely, due to the sign of T_x, T_y, T_z, we see that for a fixed time t, the temperature is decreasing as we move away from the origin and towards the boundary of the cylinder, which makes sense because the water at the edge of the pot should be cooler than the water in the middle.

Figure 3.2: A pot of water with heat being applied from the bottom.

Definition 3.2 (Gradient vector)**.** Let $E \subseteq \mathbb{R}^n$ be an open set, let $f\colon E \to \mathbb{R}$ be a function and suppose all partial derivatives $\frac{\partial f}{\partial x_1}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a})$ of f at the point $\mathbf{a} \in E$ exist. Then

$$\nabla f(\mathbf{a}) = \operatorname{grad} f(\mathbf{a}) := \left(\frac{\partial f}{\partial x_n}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right) \in \mathbb{R}^{1\times n},$$

is called *the gradient of f at* $\mathbf{a}$. If at least one of the partial derivatives $\frac{\partial f}{\partial x_1}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a})$ of f at the point $\mathbf{a}$ does not exist then we say that the gradient of f at $\mathbf{a}$ does not exist.

Remark 3.2. The gradient $\nabla f(\mathbf{a})$ can also be written as a linear combination using the canonical vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$,

$$\nabla f(\mathbf{a}) = \sum_{k=1}^{n} D_k f(\mathbf{a}) \mathbf{e}_k^\top.$$

Therefore $D_k f(\mathbf{a}) = \frac{\partial f}{\partial x_k}(\mathbf{a}) = \langle \nabla f(\mathbf{a}), \mathbf{e}_k \rangle$ for all $k = 1, 2, \ldots, n$.

3.2 Directional Derivatives

Definition 3.3 (Directional derivatives)**.** Let $E \subseteq \mathbb{R}^n$ be an open set, $f\colon E \to \mathbb{R}$ a real-valued function, and $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. The *directional derivative* of f along the vector

$\mathbf{v}$ at the point $\mathbf{a} \in E$ is defined as

$$\nabla_{\mathbf{v}} f(\mathbf{a}) = \lim_{t \to 0} \frac{f(\mathbf{a} + t\mathbf{v}) - f(\mathbf{a})}{t}$$

wherever this limit exists. If this limit does not exist then we say that the directional derivative of f along $\mathbf{v}$ at the point $\mathbf{a}$ does not exist. When $\mathbf{v}$ is a unit vector (which means $\|\mathbf{v}\|_2 = 1$), it is also called the *derivative in the direction* $\mathbf{v}$.

Note that the partial derivative with respect to the variable x_k coincides with the directional derivative along the vector $\mathbf{e}_k$, that is,

$$\frac{\partial f}{\partial x_k}(\mathbf{a}) = \nabla_{\mathbf{e}_k} f(\mathbf{a}).$$

Many of the familiar properties of the ordinary derivative hold for the directional derivative. In particular, if $\nabla_{\mathbf{v}} f(\mathbf{a})$ and $\nabla_{\mathbf{v}} g(\mathbf{a})$ exist then

1. **Linearity**: For all $\alpha, \beta \in \mathbb{R}$ we have

$$\nabla_{\mathbf{v}}(\alpha f + \beta g)(\mathbf{a}) \;=\; \alpha(\nabla_{\mathbf{v}} f(\mathbf{a})) + \beta(\nabla_{\mathbf{v}} g(\mathbf{a})).$$

2. **Product rule** (or **Leibniz's rule**):

$$\nabla_{\mathbf{v}}(f \cdot g)(\mathbf{a}) \;=\; g(\mathbf{a}) \cdot \nabla_{\mathbf{v}} f(\mathbf{a}) \;+\; f(\mathbf{a}) \cdot \nabla_{\mathbf{v}} g(\mathbf{a}).$$

3. **Quotient rule**: If $g(\mathbf{a}) \neq 0$ then

$$\nabla_{\mathbf{v}}\left(\frac{f}{g}\right)(\mathbf{a}) \;=\; \frac{g(\mathbf{a}) \cdot \nabla_{\mathbf{v}} f(\mathbf{a}) \;-\; f(\mathbf{a}) \cdot \nabla_{\mathbf{v}} g(\mathbf{a})}{g(\mathbf{a})^2}.$$

3.3 Differentiability at a Point

Recall from linear algebra that a *linear map* from $\mathbb{R}^n$ to $\mathbb{R}$ is a function $L\colon \mathbb{R}^n \to \mathbb{R}$ that satisfies linearity, meaning it preserves addition and scalar multiplication: for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all $\alpha, \beta \in \mathbb{R}$, we have

$$L(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha L(\mathbf{x}) + \beta L(\mathbf{y}).$$

Note that any linear map L can always be represented as $L(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$, where $\mathbf{w} \in \mathbb{R}^n$ is a fixed vector and $\langle ., . \rangle$ denotes the standard inner product on $\mathbb{R}^n$ defined in (1.1).

Definition 3.4 (Differentiability at a point)**.** Let E be a non-empty open subset of $\mathbb{R}^n$. A function $f\colon E \to \mathbb{R}$ is *differentiable* at the point $\mathbf{a} \in E$ if there exists a linear map $L_{\mathbf{a}}\colon \mathbb{R}^n \to \mathbb{R}$ such that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\left| f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - L_{\mathbf{a}}(\mathbf{h}) \right|}{\|\mathbf{h}\|_2} = 0.$$

In this case, the linear map $L_{\mathbf{a}}\colon \mathbb{R}^n \to \mathbb{R}$ is called the **differential** of f at the point $\mathbf{a}$.

Theorem 3.1 (Fundamental theorem)**.** *Suppose $f\colon E \to \mathbb{R}$ is a function defined on* *a set $E \subseteq \mathbb{R}^n$, and $\mathbf{a}$ is an interior point of E. If f is differentiable at $\mathbf{a}$ then the following statements hold:*

(i) f is continuous at $\mathbf{a}$.

(ii) All partial derivatives of f at the point $\mathbf{a}$ exist, the gradient vector $\nabla f(\mathbf{a})$ of f at the point $\mathbf{a}$ exists, and the differential $L_{\mathbf{a}} \colon \mathbb{R}^n \to \mathbb{R}$ of f at the point $\mathbf{a}$ is the same as scalar multiplication by the gradient vector, i.e.,

$$L_{\mathbf{a}}(\mathbf{v}) = \nabla f(\mathbf{a}) \cdot \mathbf{v}, \qquad \forall \mathbf{v} \in \mathbb{R}^n.$$

(iii) All directional derivatives of f at the point $\mathbf{a}$ exist and are given by

$$\nabla_{\mathbf{v}} f(\mathbf{a}) = L_{\mathbf{a}}(\mathbf{v}) = \nabla f(\mathbf{a}) \cdot \mathbf{v}, \qquad \forall \mathbf{v} \in \mathbb{R}^n.$$

(iv) For all $\mathbf{x} \in E$ we have

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + r_1(\mathbf{x}),$$

where r_1 is an "error" term satisfying

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{r_1(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|_2} = 0.$$

The function

$$t(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

*is called the **linearization** (or **linear approximation**) of f at the point $\mathbf{a}$.*

(v) The function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ increases most rapidly in the direction ∇f, and decreases most rapidly in the direction $-\nabla f$. Any vector $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ orthogonal to ∇f is a direction of zero change.

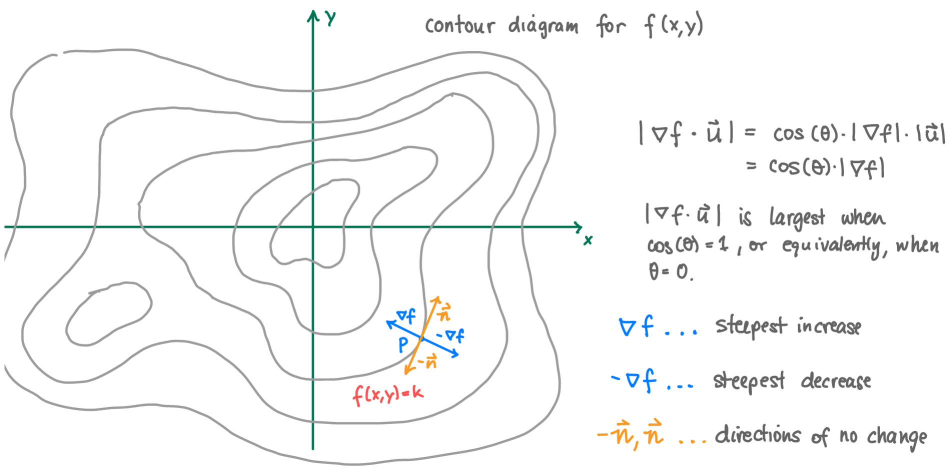


Figure 3.3: The gradient vector ∇f gives the direction of steepest incline, while the rate of change in the direction of the contour lines equals 0.

Remark 3.3. The gradient is perpendicular to the level sets of a function.

Theorem 3.2 (Sufficient conditions for differentiability)**.** *Let $E \subseteq \mathbb{R}^n$, $f\colon E \to \mathbb{R}$, and suppose* $\mathbf{a}$ *is an interior point of E. If there exists $\delta > 0$ such that every partial derivative $\frac{\partial f}{\partial x_k}$ of f exists at every point in the open ball $B(\mathbf{a}, \delta)$ and $\frac{\partial f}{\partial x_k}(x_1, \ldots, x_k)$ is a continuous function at the point* $\mathbf{a}$*, then f is differentiable at the point* $\mathbf{a}$*.*

Example 3.2. Consider $n = 2, E = \mathbb{R}^2, f\colon \mathbb{R}^2 \to \mathbb{R}, f(x, y) = x^2 - y^2$. We have:

$$\frac{\partial f}{\partial x}(x, y) = 2x,$$
$$\frac{\partial f}{\partial y}(x, y) = -2y,$$
$$\nabla f(x, y) = (2x, -2y).$$

Example 3.3. Let $E = \{(x, y) \in \mathbb{R}^2 : x > 0\}$ and $f(x, y) = e^{y \log x}$. Then

$$\frac{\partial f}{\partial x}(x, y) = \frac{y e^{y \log x}}{x},$$
$$\frac{\partial f}{\partial y}(x, y) = e^{y \log x} \cdot \log x,$$
$$\nabla f(x, y) = \left(\frac{y e^{y \log x}}{x}, e^{y \log x} \cdot \log x \right).$$

3.4 Tangent (Hyper)Planes

Recall that a straight line is called a *tangent line* to the curve $y = f(x)$ at a point $x = a$ if the line passes through the point $(a, f(a))$ on the curve and has slope $f'(a)$, where $f'(x)$ is the 1^{st} derivative of f. The equation of the tangent line is then given by

$$y = f(a) + f'(a)(x - a).$$

The equation of the tangent line is closely related to Taylor's theorem, which says that the 1^{st}-order Taylor expansion of f is given by

$$f(x) = \underbrace{\overbrace{f(a) + f'(a)(x - a)}^{\textit{linear approximation}} + \overbrace{r_1(x)}^{\textit{remainder}}}_{1^{\text{st}}\textit{-order expansion}}$$

where $r_1(x)$ is an "error" term that satisfies $\lim_{x \to a} \frac{r_1(x)}{|x-a|} = 0$.

A similar concept applies to multivariate functions in n-dimensional Euclidean space. As we have seen (cf. part (iv) of Theorem 3.1) if $f(x_1, \ldots, x_n)$ is a function in

n variables that is differentiable at a point $\mathbf{a} \in \mathbb{R}^n$ then

$$f(\mathbf{x}) = f(\mathbf{a}) + L_{\mathbf{a}}(\mathbf{x}-\mathbf{a}) + r_1(\mathbf{x}) = \underbrace{\overbrace{f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a})}^{\textit{linear approximation}} + \overbrace{r_1(\mathbf{x})}^{\textit{remainder}}}_{1^{\text{st}}\textit{-order expansion}} \tag{3.2}$$

where $r_1(\mathbf{x})$ is an "error" term satisfying $\lim_{\mathbf{x}\to\mathbf{a}} \frac{r_1(\mathbf{x})}{\|\mathbf{x}-\mathbf{a}\|_2} = 0$.

Definition 3.5 (Tangent hyperplane)**.** Let $E \subseteq \mathbb{R}^n$ and $f\colon E \to \mathbb{R}$, and assume that $\mathbf{a}$ is an interior point of E. Suppose f is differentiable at $\mathbf{a}$, and consider the linear approximation of f at $\mathbf{a}$ given by

$$t(\mathbf{x}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a}).$$

The graph of $t(\mathbf{x})$ is called the *tangent hyperplane* of f at $\mathbf{a}$. That is, the tangent hyperplane consists of all points $(x_1, \dots, x_n, x_{n+1}) \in \mathbb{R}^{n+1}$ satisfying the equation

$$x_{n+1} = t(x_1, \dots, x_n).$$

This equation is commonly referred to as the *equation of the tangent hyperplane.*

When $n = 1$, the tangent hyperplane is the same as the tangent line, and when $n = 2$ the tangent hyperplane is usually just called the *tangent plane* (see Fig. 3.4).

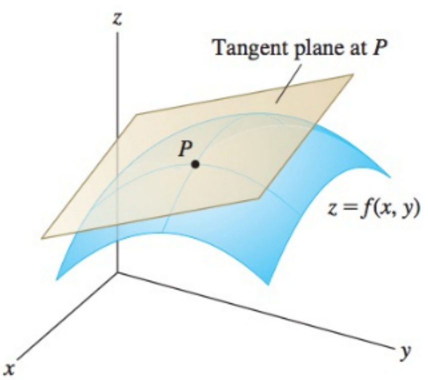


Figure 3.4: Tangent plane to a function $z = f(x, y)$ at $P = (x_0, y_0, f(x_0, y_0))$.

Example 3.4. Let us find the equation of the tangent plane to the elliptic paraboloid

$$z = 2x^2 + y^2 + 1$$

at the point $(1, -1, 4)$. This elliptic paraboloid is the graph of the function $f(x, y) = 2x^2 + y^2 + 1$. The partial derivatives of f form the gradient given by

$$\nabla f(x, y) = (4x, 2y).$$

We can now write down the linear approximation of $f(x,y)$ at the point $(1,-1)$ as

$$\begin{aligned} t(x,y) &= f(1,-1) + \nabla f(1,-1) \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) = \\ &= 4 + (4,-2) \cdot \begin{pmatrix} x-1 \\ y+1 \end{pmatrix} \\ &= 4 + 4(x-1) - 2(y+1) \\ &= 4x - 2y - 2. \end{aligned}$$

Thus, the equation of the tangent plane to the elliptic paraboloid at the point $(1,-1,4)$ is

$$z = 4x - 2y - 2.$$

3.5 Functions of Class C^1

Definition 3.6 (Differentiability in a region)**.** Let $E \subseteq \mathbb{R}^n$ be an open set and $f\colon E \to \mathbb{R}$ a function on E. If f is differentiable at every point $\mathbf{a} \in E$ then we say that f is *differentiable on* E.

Definition 3.7 (Functions of Class C^1)**.** Let $E \subseteq \mathbb{R}^n$ be an open set. A function $f\colon E \to \mathbb{R}$ is said to be *of class* $C^1(E)$ if all its partial derivatives exist and are continuous at each point $\mathbf{x} \in E$.

The existence and continuity of the partial derivatives at every point in E implies the differentiability of the function at every point in E (see Theorem 3.2). It follows that any function of class $C^1(E)$ is differentiable on E.

Proposition 3.1. *Let $E \subseteq \mathbb{R}^n$ be open and $f\colon E \to \mathbb{R}$ a function of class $C^1(E)$. Then f is differentiable on E.*

Example 3.5. Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

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

We have already studied this function in Example 2.9 and Remark 2.1.

- For $(x,y) \neq (0,0)$, we can calculate the partial derivatives as

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \frac{y}{x^2+y^2} - \frac{2x^2y}{(x^2+y^2)^2} \\ \frac{\partial f}{\partial y}(x,y) &= \frac{x}{x^2+y^2} - \frac{2xy^2}{(x^2+y^2)^2}. \end{aligned}$$

- At the point $(0,0)$ we can use the definition of partial derivatives and find

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h} = \lim_{h\to 0}\frac{0}{h} = 0$$
$$\frac{\partial f}{\partial y}(0,0) = \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h} = \lim_{h\to 0}\frac{0}{h} = 0.$$

This shows that the partial derivatives $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist for every point in $\mathbb{R}^2$. Nonetheless, this function is **not** differentiable at the point $(0,0)$. Indeed, we have seen in Example 2.9 that this function is not even continuous at the point $(0,0)$, so according to part (i) of Theorem 3.1, it cannot be differentiable at that point. This example illustrates that even if a function is differentiable in every coordinate, this does not mean that it is differentiable. In conclusion, the function f is of class $C^1(\mathbb{R}^2\backslash\{(0,0)\})$.

3.6 Second Order Partial Derivatives

The partial derivatives $\frac{\partial f}{\partial x_1},\dots,\frac{\partial f}{\partial x_n}$ are also referred to as *"partial derivatives of order 1"* or *"first order partial derivatives"*. Let us now define the second order partial derivatives.

Definition 3.8 (Partial derivatives of second order)**.** Let $E\subseteq\mathbb{R}^n$ be an open set and $1\leqslant k\leqslant n$. Assume $f\colon E\to\mathbb{R}$ is a function whose partial derivative $\frac{\partial f}{\partial x_k}$ exists for every point in E. For $1\leqslant i\leqslant n$, if the partial derivative of $\frac{\partial f}{\partial x_k}$ with respect to the variable x_i at the point a exists, then we obtain a *second order partial derivative* of f with respect to x_i and x_k at $\mathbf{a}$ denoted by $\frac{\partial^2 f}{\partial x_i\partial x_k}(a)$. If this derivative exists for every $\mathbf{a}\in E$, it defines a function $\frac{\partial^2 f}{\partial x_i\partial x_k}: E\to\mathbb{R}$.

If $i=k$, then it is also common to write $\frac{\partial^2 f}{\partial x_i^2}$ instead of $\frac{\partial^2 f}{\partial x_i\partial x_i}$. If $i\neq k$, then there are generally two mixed second-order partial derivatives:

$$\frac{\partial^2 f}{\partial x_i\partial x_k} \quad\text{and}\quad \frac{\partial^2 f}{\partial x_k\partial x_i}.$$

These derivatives are not necessarily equal since the order of differentiation can affect the result. However, as the following theorem states, they are equal if an additional continuity assumption is satisfied.

Theorem 3.3 (Schwarz's theorem)**.** *Let $E\subseteq\mathbb{R}^n$ be open and let $f\colon E\to\mathbb{R}$ be a function defined on E. For any point $\mathbf{a}\in E$ and indices $i,k\in\{1,\dots,n\}$, suppose the mixed partial derivatives $\frac{\partial^2 f}{\partial x_i\partial x_k}$ and $\frac{\partial^2 f}{\partial x_k\partial x_i}$ exist in E and are continuous at $\mathbf{a}$. Then, $\frac{\partial^2 f}{\partial x_i\partial x_k}(\mathbf{a})=\frac{\partial^2 f}{\partial x_k\partial x_i}(\mathbf{a})$.*

Definition 3.9. The $n \times n$ matrix

$$\operatorname{Hess}(f)(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{pmatrix}$$

is called the *Hessian matrix* of f at the point $\mathbf{a}$, written $\operatorname{Hess}(f)(\mathbf{a})$.

If all the partial derivatives of order 2 exist and are continuous at $\mathbf{a}$ then by Schwarz's theorem the Hessian matrix is a symmetric matrix, i.e., $\operatorname{Hess}(f)(\mathbf{a}) = \operatorname{Hess}(f)(\mathbf{a})^T$. In this case we can use the Hessian matrix to form the *second order expansion* of a differentiable function, given by

$$f(\mathbf{x}) = \underbrace{\overbrace{\overbrace{f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a})}^{\text{linear approximation}} + \frac{1}{2}(\mathbf{x}-\mathbf{a})^T \cdot \operatorname{Hess}(f)(\mathbf{a}) \cdot (\mathbf{x}-\mathbf{a})}^{\text{quadratic approximation}} + \overbrace{r_2(\mathbf{x})}^{\text{remainder}}}_{2^{\text{nd}}\text{-order expansion}} \tag{3.3}$$

where $r_2(\mathbf{x})$ is an "error" term satisfying $\lim_{\mathbf{x}\to\mathbf{a}} \frac{r_2(\mathbf{x})}{\|\mathbf{x}-\mathbf{a}\|_2^2} = 0$.

The quadratic approximation is a polynomial of degree 2 in n variables called the *Taylor polynomial of order* 2 *at the point* $\mathbf{a}$ and it is usually denoted by $P_2(x, y)$.

Example 3.6. Let us find the Taylor polynomial of order 2 for the function $f(x, y) = \sin(2x + y) + 3\cos(x + y)$ at the point $(0, 0)$. Recall the formula for computing the quadratic approximation of a function in two variables at the point $(0, 0)$ is

$$P_2(x, y) = f(0, 0) + \nabla f(0, 0) \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2}(x, y) \cdot \operatorname{Hess}(f)(0, 0) \cdot \begin{pmatrix} x \\ y \end{pmatrix}.$$

To use this formula, we have to find the gradient vector and the Hessian matrix first. We have

$$\nabla f(x, y) = (2\cos(2x + y) - 3\sin(x + y), \cos(2x + y) - 3\sin(x + y))$$

which gives

$$\nabla f(0, 0) = (2, 1).$$

Moreover,

$$\operatorname{Hess}(f)(x, y) = \begin{pmatrix} -4\sin(2x + y) - 3\cos(x + y) & -2\sin(2x + y) - 3\cos(x + y) \\ -2\sin(2x + y) - 3\cos(x + y) & -\sin(2x + y) - 3\cos(x + y) \end{pmatrix}$$

and hence

$$\operatorname{Hess}(f)(0, 0) = \begin{pmatrix} -3 & -3 \\ -3 & -3 \end{pmatrix}.$$

It follows that

$$\begin{aligned} P_2(x,y) &= 3 + (2,1)\cdot\begin{pmatrix} x \\ y \end{pmatrix} + \frac{1}{2}(x,y)\cdot\begin{pmatrix} -3 & -3 \\ -3 & -3 \end{pmatrix}\cdot\begin{pmatrix} x \\ y \end{pmatrix} \\ &= 3 + 2x + y - \frac{3}{2}x^2 - 3xy - \frac{3}{2}y^2. \end{aligned}$$

This is a degree 2 polynomial in 2 variables.

3.7 Higher Order Partial Derivatives

Definition 3.10 (Partial derivatives of higher orders)**.** Consider a function $f\colon E \to \mathbb{R}$ defined on an open set $E \subseteq \mathbb{R}^n$. For a sequence of indices $i_1, \ldots, i_p$ with each $i_j \in \{1, \ldots, n\}$ and for $p \geqslant 3$, assume that the $(p-1)$-th order partial derivative of f, denoted as $\frac{\partial^{p-1} f}{\partial x_{i_1} \ldots \partial x_{i_{p-1}}}$, exists in E. Then, the p-th order partial derivative of f with respect to these indices, if it exists, is given by:

$$\frac{\partial^p f}{\partial x_{i_p} \ldots \partial x_{i_1}} = \frac{\partial}{\partial x_{i_p}}\left(\frac{\partial^{p-1} f}{\partial x_{i_1} \ldots \partial x_{i_{p-1}}}\right).$$

This derivative is denoted as $\frac{\partial f}{\partial x_{i_p} \ldots \partial x_{i_1}}(\mathbf{a})$ for any point $\mathbf{a} \in E$. If such a derivative exists for every $\mathbf{a} \in E$, it defines a function $\frac{\partial^p f}{\partial x_{i_p} \ldots \partial x_{i_1}} : E \to \mathbb{R}$.

Example 3.7. Consider a function $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ defined by $f(x,y) = x^3y^2$. We calculate its higher-order partial derivatives as follows:

$$\begin{aligned} &\frac{\partial f}{\partial x}(x,y) = 3x^2y^2, \\ &\frac{\partial^2 f}{\partial x^2}(x,y) = \frac{\partial}{\partial x}(3x^2y^2) = 6xy^2, \\ &\frac{\partial^2 f}{\partial y \partial x}(x,y) = \frac{\partial}{\partial y}(3x^2y^2) = 6x^2y, \\ &\frac{\partial^3 f}{\partial y \partial x^2}(x,y) = \frac{\partial}{\partial y}(6xy^2) = 12xy, \\ &\frac{\partial^3 f}{\partial x^3}(x,y) = \frac{\partial}{\partial x}(6xy^2) = 6y^2. \end{aligned}$$

This illustrates the computation of first, second, and third-order partial derivatives for a function of two variables.

Remark 3.4. Explicit computations also give $\frac{\partial^2 f}{\partial x \partial y}(x,y) = 6x^2y = \frac{\partial^2 f}{\partial y \partial x}(x,y)$ and $\frac{\partial^3 f}{\partial x \partial y \partial x}(x,y) = 12xy = \frac{\partial^3 f}{\partial y \partial x^2}(x,y)$, demonstrating the symmetry in mixed partial derivatives.

3.8 Functions of class C^p

Definition 3.11 (Functions of class C^p)**.** Let E be an open subset of $\mathbb{R}^n$ and p a positive integer. A function $f: E \to \mathbb{R}$ is said to be *of class* $C^p(E)$ if all its partial derivatives of order p exist and are continuous at every point in E.
A function $f: E \to \mathbb{R}$ is said to be of class $C^\infty(E)$ if, for every integer $p > 0$, it is of class $C^p(E)$.

Proposition 3.2. *If $f: E \to \mathbb{R}$ is a function of class $C^p(E)$, then it is also of class $C^k(E)$ for all $0 < k \leqslant p$.*

Example 3.8. Consider the function $f: \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x\sin(xy)$. Then, for every $(x, y) \in \mathbb{R}^2$, we have:

$$\frac{\partial f}{\partial x}(x,y) = \sin(xy) + xy\cos(xy),$$
$$\frac{\partial f}{\partial y}(x,y) = x^2\cos(xy),$$
$$\frac{\partial^2 f}{\partial x^2}(x,y) = 2y\cos(xy) - xy^2\sin(xy),$$
$$\frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{\partial^2 f}{\partial y \partial x}(x,y) = 2x\cos(xy) - x^2 y\sin(xy),$$
$$\frac{\partial^2 f}{\partial y^2}(x,y) = -x^3\sin(xy).$$

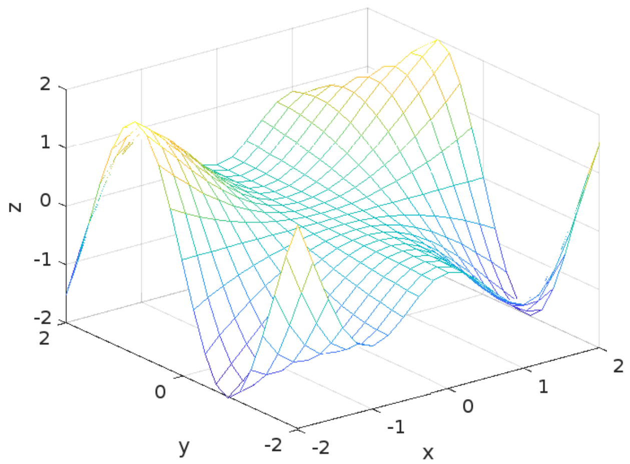


Figure 3.5: $f(x, y) = x\sin(xy)$

The following is a corollary of Schwarz's theorem.

Corollary 3.1. *Let $f: E \to \mathbb{R}$ be a function of class $C^p(E)$ and let k be an integer between 1 and p. If two ordered k-tuples $(i_1, \cdots, i_k)$ and $(j_1, \cdots, j_k)$ are equal up to* *a permutation, then, for any element* $\mathbf{a} = (a_1, \ldots, a_n)$ *of* E*, we have*

$$\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a_1, \cdots, a_n) = \frac{\partial^k f}{\partial x_{j_1} \cdots \partial x_{j_k}}(a_1, \cdots, a_n).$$

3.9 Taylor's Theorem for Multivariable Functions

The following is a special (but often very useful) case of Taylor's theorem for multivariate functions.

Theorem 3.4 (Taylor's Formula – special case)**.** *Let* $E \subseteq \mathbb{R}^n$ *be open and* $f\colon E \to \mathbb{R}$ *a function of class* $C^{p+1}(E)$*. Then for every* $\mathbf{a} \in E$ *there exists a real number* $\delta > 0$ *such that* $B(\mathbf{a}, 2\delta) \subseteq E$ *and, for every element* $\mathbf{x} \in B(\mathbf{a}, \delta)$*, one can associate a number* $0 < \theta < 1$ *so that the following equality (known as Taylor's formula) holds:*

$$f(\mathbf{x}) = F(0) + F'(0) + \ldots + F^{(p)}(0)\frac{1}{p!} + F^{(p+1)}(\theta)\frac{1}{(p+1)!},$$

where $F\colon (-2, 2) \to \mathbb{R}$ *is the function defined by* $F(t) = f(\mathbf{a} + t(\mathbf{x} - \mathbf{a}))$*.*

To state Taylor's theorem for multivariate functions in full generality, we first need to introduce the multi-index notation. Given an n-tuple of non-negative integers $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and a point $\mathbf{x} \in \mathbb{R}^n$, let

$$|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_n, \quad \boldsymbol{\alpha}! = \alpha_1! \cdots \alpha_n!, \quad \mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

(Recall that by convention $0! = 1$.) For example, if $n = 3$ and $\boldsymbol{\alpha} = (1, 0, 4)$ then we have $|\boldsymbol{\alpha}| = 1 + 0 + 4 = 5$, and $\boldsymbol{\alpha}! = 1! \cdot 0! \cdot 4! = 24$, and $(x_1, x_2, x_3)^\alpha = x_1 x_3^4$. Given a function $f\colon E \to \mathbb{R}$ of class $C^k(E)$ and an n-tuple of non-negative integers $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ with $|\boldsymbol{\alpha}| \leqslant k$ then we write

$$D^{\boldsymbol{\alpha}} f = \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

Since f is of class $C^k(E)$, all its k-th order partial derivatives exist and are continuous and, by Schwarz's theorem, one can change the order of mixed derivatives. This ensures that as long as $|\boldsymbol{\alpha}| \leqslant k$ the above notation is well-defined and unambiguous.

Theorem 3.5 (Multivariate version of Taylor's theorem)**.** *Let* $k \in \mathbb{N}$*. Suppose* $E \subseteq \mathbb{R}^n$ *is open and* $f\colon E \to \mathbb{R}$ *is a function of class* $C^k(E)$*. Then*

$$f(\mathbf{x}) = \underbrace{\overbrace{\sum_{|\boldsymbol{\alpha}| \leqslant k} \frac{D^{\boldsymbol{\alpha}} f(\mathbf{a})}{\boldsymbol{\alpha}!}(\mathbf{x} - \mathbf{a})^{\boldsymbol{\alpha}}}^{k^{\text{th}}\text{-order approximation}} + \overbrace{r_k(\mathbf{x})}^{\text{remainder}}}_{k^{\text{th}}\text{-order expansion}} \tag{3.4}$$

where the sum is taken over all n*-tuples of non-negative integers* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ *with* $|\boldsymbol{\alpha}| \leqslant k$ *and* $r_k(\mathbf{x})$ *is an "error" term satisfying* $\lim_{\mathbf{x} \to \mathbf{a}} \frac{r_k(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|_2^k} = 0$.

Note that if $k = 1$ then formula (3.4) is the same as (3.2) and if $k = 2$ then formula (3.4) is the same as (3.3).