1

MULTIVARIABLE CALCULUS

IN THIS CHAPTER we consider functions mapping $\mathbb{R}^m$ into $\mathbb{R}^n$, and we define what we mean by the derivative of such a function. It is important to be familiar with the idea that the derivative at a point a of a map between open sets of (normed) vector spaces is a *linear transformation* between the vector spaces (in this chapter the linear transformation is represented as a $n \times m$ matrix).

This chapter is based on Spivak (1965, Chapters 1 & 2) and Munkres (1991, Chapter 2)—one could do no better than to study theses two excellent books for multivariable calculus.

Notation

We use standard notation:

$$\begin{aligned}
\mathbb{N} &:= \{1, 2, 3, \ldots\} = \text{the set of all natural numbers,}\\
\mathbb{Z} &:= \{0, \pm 1, \pm 2, \ldots\} = \text{the set of all integers,}\\
\mathbb{Q} &:= \left\{\frac{n}{m} : n, m \in \mathbb{Z} \text{ and } m \neq 0\right\} = \text{the set of all rational numbers,}\\
\mathbb{R} &:= \text{the set of all real numbers.}
\end{aligned}$$

We also define

$$\mathbb{R}_+ := \{x \in \mathbb{R} : x \geqslant 0\} \quad \text{and} \quad \mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}.$$

1.1 Functions on Euclidean Space

Norm, Inner Product and Metric

▶ **Definition 1.1** (Euclidean n-space) *Euclidean n-space* $\mathbb{R}^n$ is defined as the set of all n-tuples $(x_1, \ldots, x_n)$ of real numbers x_i:

$$\mathbb{R}^n := \left\{(x_1, \ldots, x_n) : \text{each } x_i \in \mathbb{R}\right\}.$$

An element of $\mathbb{R}^n$ is often called a point in $\mathbb{R}^n$, and $\mathbb{R}^1$, $\mathbb{R}^2$, $\mathbb{R}^3$ are often called the *line*, the *plane*, and *space*, respectively.

If $\boldsymbol{x}$ denotes an element of $\mathbb{R}^n$, then $\boldsymbol{x}$ is an n-tuple of numbers, the i^{th} one of which is denoted x_i; thus, we can write

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

A point in $\mathbb{R}^n$ is frequently also called a *vector in* $\mathbb{R}^n$, because $\mathbb{R}^n$, with

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n) \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$

and

$$\alpha\boldsymbol{x} = (\alpha x_1, \ldots, \alpha x_n), \quad \alpha \in \mathbb{R} \text{ and } \boldsymbol{x} \in \mathbb{R}^n,$$

as operations, is a *vector space*.

To obtain the full geometric structure of $\mathbb{R}^n$, we introduce three structures on $\mathbb{R}^n$: the *Euclidean norm*, *inner product* and *metric*.

▶ **Definition 1.2** (Norm) In $\mathbb{R}^n$, the length of a vector $\boldsymbol{x} \in \mathbb{R}^n$, usually called the *norm* $\|\boldsymbol{x}\|$ of $\boldsymbol{x}$, is defined by

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + \cdots x_n^2}.$$

Remark 1.3 The norm $\|\cdot\|$ satisfies the following properties: for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$,

- $\|\boldsymbol{x}\| \geqslant 0$,
- $\|\boldsymbol{x}\| = 0$ iff[1] $\boldsymbol{x} = \boldsymbol{0}$,
- $\|\alpha\boldsymbol{x}\| = |\alpha| \cdot \|\boldsymbol{x}\|$,
- $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ (Triangle inequality).

[1] "iff" is the abbreviation of "if and only if".

Exercise 1.4 Prove that $\big|\|x\| - \|y\|\big| \leqslant \|x - y\|$ for any two vectors $x, y \in \mathbb{R}^n$ (use the triangle inequality).

▶ **Definition 1.5** (Inner Product) Given $x, y \in \mathbb{R}^n$, the *inner product* of the vectors x and y, denoted $x \cdot y$ or $\langle x, y \rangle$, is defined as

$$x \cdot y = \sum_{i=1}^{n} x_i y_i .$$

Remark 1.6 The norm and the inner product are related through the following identity:

$$\|x\| = \sqrt{x \cdot x}.$$

■ **Theorem 1.7** (Cauchy-Schwartz Inequality) *For any $x, y \in \mathbb{R}^n$ we have*

$$|x \cdot y| \leqslant \|x\| \, \|y\|.$$

Proof. We assume that $x \neq 0$; for otherwise the proof is trivial. For every $a \in \mathbb{R}$, we have

$$0 \leqslant \|ax + y\|^2 = a^2\|x\|^2 + 2a(x \cdot y) + \|y\|^2.$$

In particular, let $a = -(x \cdot y)/\|x\|^2$. Then, from the above display, we get the desired result. □

Exercise 1.8 Prove the triangle inequality (use the Cauchy-Schwartz Inequality). Show it holds with equality iff one of the vector is a nonnegative scalar multiple of the other.

▶ **Definition 1.9** (Metric) The *distance* $d(x, y)$ between two vectors $x, y \in \mathbb{R}^n$ is given by

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

The distance function d is called a *metric*.

◇ **Example 1.10** In $\mathbb{R}^2$, choose two points $x^1 = (x_1^1, x_2^1)$ and $x^2 = (x_1^2, x_2^2)$ with $x_1^2 - x_1^1 = a$ and $x_2^2 - x_2^1 = b$. Then Pythagoras tells us that (Figure 1.1)

$$d(x^1, x^2) = \sqrt{a^2 + b^2} = \sqrt{\left(x_1^2 - x_1^1\right)^2 + \left(x_2^2 - x_2^1\right)^2}.$$

Remark 1.11 The metric is related to the norm $\| \cdot \|$ through the identity

$$d(x, y) = \|x - y\|.$$

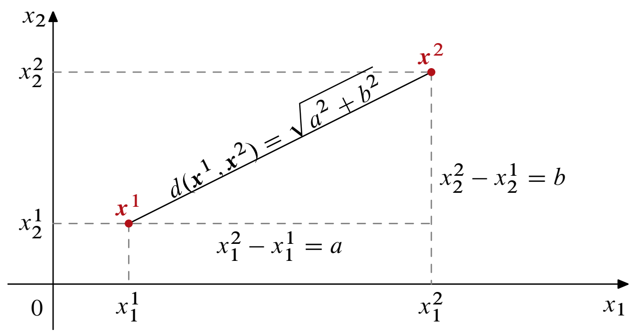


Figure 1.1: Distance in the plane.

Subsets of $\mathbb{R}^n$

▶ **Definition 1.12** (Open Ball) Let $\boldsymbol{x} \in \mathbb{R}^n$ and $r > 0$. The *open ball* $\mathbb{B}(\boldsymbol{x}; r)$ with center $\boldsymbol{x}$ and radius r is given by

$$\mathbb{B}(\boldsymbol{x}; r) := \left\{\boldsymbol{y} \in \mathbb{R}^n : d(\boldsymbol{x}, \boldsymbol{y}) < r\right\}.$$

▶ **Definition 1.13** (Interior) Let $S \subset \mathbb{R}^n$. A point $\boldsymbol{x} \in S$ is called an *interior point* of S if there is some $r > 0$ such that $\mathbb{B}(\boldsymbol{x}; r) \subset S$. The set of all interior points of S is called its *interior* and is denoted S°.

▶ **Definition 1.14** Let $S \subset \mathbb{R}^n$.

- S is *open* if for every $\boldsymbol{x} \in S$ there exists $r > 0$ such that $\mathbb{B}(\boldsymbol{x}; r) \subset S$.
- S is *closed* if its complement $\mathbb{R}^n \setminus S$ is open.
- S is *bounded* if there exists $r > 0$ such that $S \subset \mathbb{B}(\boldsymbol{0}; r)$.
- S is *compact* if (and only if) it is closed and bounded (Heine-Borel Theorem).[2]

◇ **Example 1.15** On $\mathbb{R}$, the interval $(0, 1)$ is open, the interval $[0, 1]$ is closed. Both $(0, 1)$ and $[0, 1]$ are bounded, and $[0, 1]$ is compact. However, the interval $(0, 1]$ is neither open nor closed. But $\mathbb{R}$ is both open and closed.

[2] This definition does not work for more general metric spaces. See Willard (2004) for details.

Limit and Continuity

FUNCTIONS A *function* from $\mathbb{R}^m$ to $\mathbb{R}^n$ (sometimes called a vector-valued function of m variables) is a rule which associates to each point in $\mathbb{R}^m$ some point in $\mathbb{R}^n$. We write

$$f : \mathbb{R}^m \to \mathbb{R}^n$$

to indicate that $f(\boldsymbol{x}) \in \mathbb{R}^n$ is defined for $\boldsymbol{x} \in \mathbb{R}^m$.

The notation $f : A \to \mathbb{R}^n$ indicates that $f(\boldsymbol{x})$ is defined only for $\boldsymbol{x}$ in the set A, which is called the *domain* of f. If $B \subset A$, we define $f(B)$ as the set of all $f(\boldsymbol{x})$ for $\boldsymbol{x} \in B$:

$$f(B) := \{f(\boldsymbol{x}) : \boldsymbol{x} \in B\}.$$

If $C \subset \mathbb{R}^n$ we define

$$f^{-1}(C) := \{\boldsymbol{x} \in A : f(\boldsymbol{x}) \in C\}.$$

The notation $f : A \to B$ indicates that $f(A) \subset B$. The *graph* of $f : A \to B$ is the set of all pairs $(a, b) \in A \times B$ such that $b = f(a)$.

A function $f : A \to \mathbb{R}^n$ determines n *component functions* $f_1, \ldots, f_n : A \to \mathbb{R}$ by

$$f(\boldsymbol{x}) = \big(f_1(\boldsymbol{x}), \ldots, f_n(\boldsymbol{x})\big).$$

SEQUENCES A *sequence* is a function that assigns to each natural number $n \in \mathbb{N}$ a vector or point $\boldsymbol{x}_n \in \mathbb{R}^n$. We usually write the sequences as $(\boldsymbol{x}_n)_{n=1}^{\infty}$ or $(\boldsymbol{x}_n)$.

◇ **Example 1.16** Examples of sequences in $\mathbb{R}^2$ are

(a) $(\boldsymbol{x}_n) = ((n, n))$.

(b) $(\boldsymbol{x}_n) = ((\cos \frac{n\pi}{2}, \sin \frac{n\pi}{2}))$.

(c) $(\boldsymbol{x}_n) = (((-1)^n/2^n, 1/2^n))$.

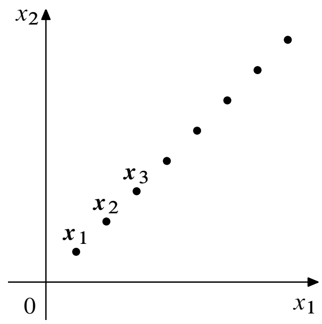
(d) $(\boldsymbol{x}_n) = (((-1)^n - 1/n, (-1)^n - 1/n))$.

See Figure 1.2.

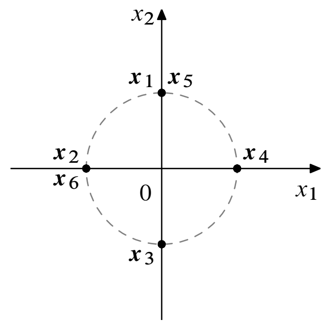
▶ **Definition 1.17** (Limit) A sequence $(\boldsymbol{x}_n)$ is said to have a *limit* $\boldsymbol{x}$ or to *converge* to $\boldsymbol{x}$ if for every $\varepsilon > 0$ there is $N_\varepsilon \in \mathbb{N}$ such that whenever $n > N_\varepsilon$, we have $\boldsymbol{x}_n \in \mathbb{B}(\boldsymbol{x}; \varepsilon)$. We write

$$\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x} \quad \text{or} \quad \boldsymbol{x}_n \to \boldsymbol{x}.$$

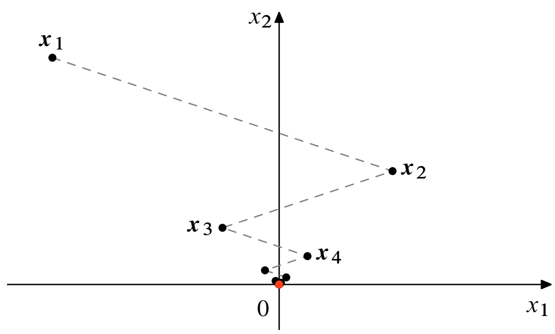
◇ **Example 1.18** In Example 1.16, the sequences (a), (b) and (d) do not converge, while the sequence (c) converges to $(0, 0)$.



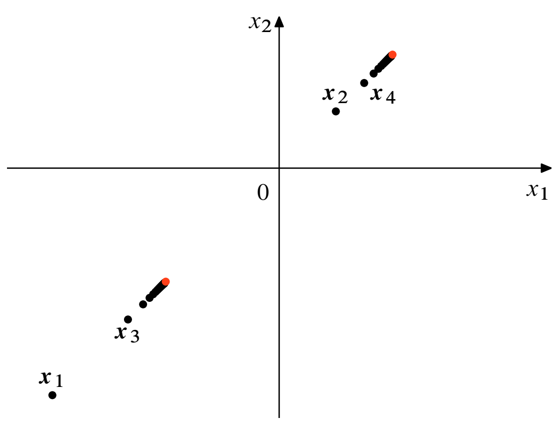
(a) $(\boldsymbol{x}_n) = \big((n, n)\big)$



(b) $(\boldsymbol{x}_n) = \Big((\cos \frac{n\pi}{2}, \sin \frac{n\pi}{2})\Big)$

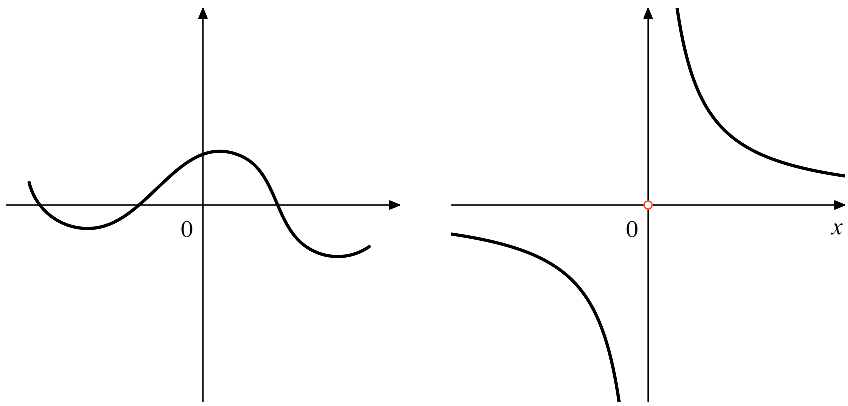


(c) $(\boldsymbol{x}_n) = \big\{((-1)^n/2^n, 1/2^n)\big\}$, which is convergent



(d) $(\boldsymbol{x}_n) = \big(((-1)^n - 1/n, (-1)^n - 1/n)\big)$

Figure 1.2: Examples of sequences



(a) A continuous function.

(b) We cannot draw the graph of $1/x$ without taking our pencil off the paper.

Figure 1.3: Naive continuity.

Naive Continuity The simplest way to say that a function $f : A \to \mathbb{R}$ is *continuous* is to say that one can draw its graph without taking the pencil off the paper. For example, a function whose graph looks like in Figure 1.3(a) would be continuous in this sense (Crossley, 2005, Chapter 2).

But if we look at the function $f(x) = 1/x$, then we see that things are not so simple. The graph of this function has two parts—one part corresponding to negative x values, and the other to positive x values. The function is not defined at 0, so we certainly cannot draw both parts of this graph without taking our pencil off the paper; see Figure 1.3(b). Of course, $f(x) = 1/x$ is continuous near every point in its domain. Such a function deserves to be called continuous. So this characterization of continuity in terms of graph-sketching is too simplistic.

Rigorous Continuity The notation $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = \boldsymbol{b}$ means, as in the one-variable case, that we get $f(\boldsymbol{x})$ as close to $\boldsymbol{b}$ as desired, by choosing $\boldsymbol{x}$ sufficiently close to, but not equal to, $\boldsymbol{a}$. In mathematical terms this means that for every number $\varepsilon > 0$ there is a number $\delta > 0$ such that $\|f(\boldsymbol{x}) - \boldsymbol{b}\| < \varepsilon$ for all $\boldsymbol{x}$ in the domain of f which satisfy $0 < \|\boldsymbol{x} - \boldsymbol{a}\| < \delta$.

A function $f : A \to \mathbb{R}^n$ is called *continuous at* $\boldsymbol{a} \in A$ if $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a})$, and f is continuous if it is continuous at each $\boldsymbol{a} \in A$.

Exercise 1.19 Let

$$f(x) = \begin{cases} x & \text{if } x \neq 1 \\ 3/2 & \text{if } x = 1. \end{cases}$$

Show that $f(x)$ is not continuous at $a = 1$.

1.2 Directional Derivative and Derivative

Let us first recall how the derivative of a real-valued function of a real variable is defined. Let $A \subset \mathbb{R}$; let $f : A \to \mathbb{R}$. Suppose A contains a neighborhood of the point a, that is, there is an open ball $\mathbb{B}(a; r)$ such that $\mathbb{B}(a; r) \subset A$. We define the *derivative of f at a* by the equation

$$f'(a) = \lim_{t \to 0} \frac{f(a+t) - f(a)}{t}, \tag{1.1}$$

provided the limit exists. In this case, we say that f is *differentiable at a*. Geometrically, $f'(a)$ is the slope of the tangent line to the graph of f at the point $(a, f(a))$.

Definition 1.20 For a function $f : (a, b) \to \mathbb{R}$, and point $x_0 \in (a, b)$, if

$$\lim_{t \uparrow 0} \frac{f(x_0 + t) - f(x_0)}{t}$$

exists and is finite, we denote this limit by $f'_-(x_0)$ and call it the *left-hand derivative* of f at x_0. Similarly, we define $f'_+(x_0)$ and call it the *right-hand derivative* of g at x_0. Of course, f is differentiable at x_0 iff it has left-hand and right-hand derivatives at x_0 that are equal.

Now let $A \subset \mathbb{R}^m$, where $m > 1$; let $f : A \to \mathbb{R}^n$. Can we define the derivative of f by replacing a and t in the definition just given by points of $\mathbb{R}^m$? Certainly we cannot since we cannot divide a point of $\mathbb{R}^n$ by a point of $\mathbb{R}^m$ if $m > 1$.

Directional Derivative

The following is our first attempt at a definition of "derivative". Let $A \subset \mathbb{R}^m$ and let $f : A \to \mathbb{R}^n$. We study how f changes as we move from a point $\boldsymbol{a} \in A^\circ$ (the interior of A) along a line segment to a nearby point $\boldsymbol{a} + \boldsymbol{u}$, where $\boldsymbol{u} \neq \boldsymbol{0}$. Each point on the segment can be expressed as $\boldsymbol{a} + t\boldsymbol{u}$, where $t \in \mathbb{R}$. The vector $\boldsymbol{u}$ describes the direction of the line segment. Since $\boldsymbol{a}$ is an interior point of A, the line segment joining $\boldsymbol{a}$ to $\boldsymbol{a} + t\boldsymbol{u}$ will lie in A if t is small enough.

▶ **Definition 1.21** (Directional Derivative) Let $A \subset \mathbb{R}^m$; let $f : A \to \mathbb{R}^n$. Suppose A contains a neighborhood of $\boldsymbol{a}$. Given $\boldsymbol{u} \in \mathbb{R}^m$ with $\boldsymbol{u} \neq \boldsymbol{0}$, define

$$f'(\boldsymbol{a}; \boldsymbol{u}) = \lim_{t \to 0} \frac{f(\boldsymbol{a} + t\boldsymbol{u}) - f(\boldsymbol{a})}{t},$$

provided the limit exists. This limit is called the *directional derivative of f at $\boldsymbol{a}$ with respect to the vector $\boldsymbol{u}$*.

Remark 1.22 In calculus, one usually requires $\boldsymbol{u}$ to be a unit vector, i.e., $\|\boldsymbol{u}\| = 1$, but that is not necessary.

◇ **Example 1.23** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by the equation $f(x_1, x_2) = x_1 x_2$. The directional derivative of f at $\boldsymbol{a} = (a_1, a_2)$ with respect to the vector $\boldsymbol{u} = (u_1, u_2)$ is

$$\begin{aligned} f'(\boldsymbol{a}; \boldsymbol{u}) = \lim_{t \to 0} \frac{f(\boldsymbol{a} + t\boldsymbol{u}) - f(\boldsymbol{a})}{t} &= \lim_{t \to 0} \frac{(a_1 + tu_1)(a_2 + tu_2) - a_1 a_2}{t} \\ &= u_2 a_1 + u_1 a_2. \end{aligned}$$

◇ **Example 1.24** Suppose the directional derivative of f at $\boldsymbol{a}$ with respect to $\boldsymbol{u}$ exists. Then for $c \in \mathbb{R}$,

$$\begin{aligned} f'(\boldsymbol{a}; c\boldsymbol{u}) = \lim_{t \to 0} \frac{f(\boldsymbol{a} + tc\boldsymbol{u}) - f(\boldsymbol{a})}{t} &= c \lim_{t \to 0} \frac{f(\boldsymbol{a} + tc\boldsymbol{u}) - f(\boldsymbol{a})}{tc} \\ &= c \lim_{s \to 0} \frac{f(\boldsymbol{a} + s\boldsymbol{u}) - f(\boldsymbol{a})}{s} \\ &= c f'(\boldsymbol{a}; \boldsymbol{u}). \end{aligned}$$

Remark 1.25 Example 1.24 shows that if $\boldsymbol{u}$ and $\boldsymbol{v}$ are collinear vectors in $\mathbb{R}^m$, then $f'(\boldsymbol{a}; \boldsymbol{u})$ and $f'(\boldsymbol{a}; \boldsymbol{v})$ are collinear in $\mathbb{R}^n$.

However, directional derivative is *not* the appropriate generalization of the notion of "derivative". The main problems are:

Problem 1. Continuity does not follow from this definition of "differentiability". There exists functions such that $f'(\boldsymbol{a}; \boldsymbol{u})$ exists for all $\boldsymbol{u} \neq \boldsymbol{0}$ but are not continuous.

◇ **Example 1.26** Define $f : \mathbb{R}^2 \to \mathbb{R}$ by setting

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

We show that all directional derivatives of f exist at $\mathbf{0}$, but that f is not continuous at $\mathbf{0}$. Let $\boldsymbol{u} = (h,k) \neq \mathbf{0}$. Then

$$\frac{f(\mathbf{0}+t\boldsymbol{u})-f(\mathbf{0})}{t} = \frac{(th)^2(tk)}{\left[(th)^4+(tk)^2\right]t} = \frac{h^2k}{t^2h^4+k^2},$$

so that

$$f'(\mathbf{0};\boldsymbol{u}) = \begin{cases} h^2/k & \text{if } k \neq 0, \\ 0 & \text{if } k = 0. \end{cases}$$

However, the function f takes the value $1/2$ as each point of the parabola $y = x^2$ (except at $\mathbf{0}$), so f is not continuous at $\mathbf{0}$ since $f(\mathbf{0}) = 0$.

Problem 2. Composites of "differentiable" functions may not differentiable.

Derivative

To give the right generalization of the notion of "derivative", let us reconsider (1.1). In fact, if $f'(a)$ exists, let $R_a(t)$ denote the difference

$$R_a(t) := \begin{cases} \frac{f(a+t)-f(a)}{t} - f'(a) & \text{if } t \neq 0 \\ 0 & \text{if } t = 0. \end{cases} \tag{1.2}$$

From (1.2) we see that $\lim_{t\to 0} R_a(t) = 0$. Then we have

$$f(a+t) = f(a) + f'(a)t + R_a(t)t. \tag{1.3}$$

Note that (1.3) also holds for $t = 0$. This is called the *first-order Taylor formula* for approximating $f(a+t) - f(a)$ by $f'(a)t$. The error committed is $R_a(t)t$.[3] See Figure 1.4. It is this idea leads to the following definition:

▶ **Definition 1.27** (Differentiability) Let $A \subset \mathbb{R}^m$; let $f\colon A \to \mathbb{R}^n$. Suppose A contains a neighborhood of $\boldsymbol{a}$. We say that f is *differentiable at* $\boldsymbol{a}$ if there is an $n \times m$ matrix $\mathbf{B}_{\boldsymbol{a}}$ such that

$$f(\boldsymbol{a}+\boldsymbol{h}) = f(\boldsymbol{a}) + \mathbf{B}_{\boldsymbol{a}} \cdot \boldsymbol{h} + \|\boldsymbol{h}\|\, R_{\boldsymbol{a}}(\boldsymbol{h}),$$

where $\lim_{\boldsymbol{h}\to\mathbf{0}} R_{\boldsymbol{a}}(\boldsymbol{h}) = \mathbf{0}$. The matrix $\mathbf{B}_{\boldsymbol{a}}$, which is unique, is called the *derivative of f at $\boldsymbol{a}$*; it is denoted $\mathbf{D}f(\boldsymbol{a})$.

[3] "All science is dominated by the idea of approximation."—Bertrand Russell.

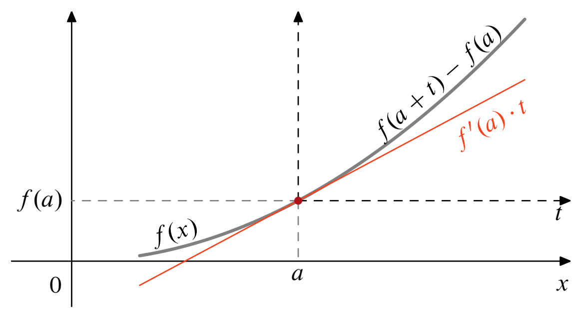


Figure 1.4: $f'(a)t$ is the linear approximation to $f(a+t) - f(a)$ at a.

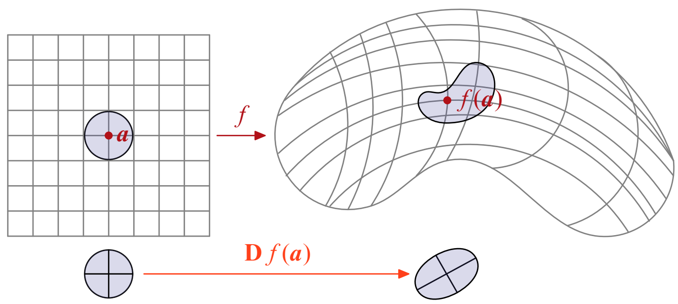


Figure 1.5: $\mathbf{D}f(\boldsymbol{a})$ is the linear part of f at $\boldsymbol{a}$.

Remark 1.28 **[1]** Notice that $\boldsymbol{h}$ is a point of $\mathbb{R}^m$ and $f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a}) - \mathbf{B}_{\boldsymbol{a}} \cdot \boldsymbol{h}$ is a point of $\mathbb{R}^n$, so the norm signs are essential.

[2] The derivative $\mathbf{D}f(\boldsymbol{a})$ depends on the point $\boldsymbol{a}$ as well as the function f. We are not saying that there exists a $\mathbf{B}$ which works for all $\boldsymbol{a}$, but that for a fixed $\boldsymbol{a}$ such a $\mathbf{B}$ exists.

[3] Here is how to visualize $\mathbf{D}f$. Take $m = n = 2$. The function $f: A \to \mathbb{R}^2$ distorts shapes nonlinearly; its derivative describes the linear part of the distortion. Circles are sent by f to wobbly ovals, but they become ellipses under $\mathbf{D}f(\boldsymbol{a})$ (here we treat $\mathbf{D}f(\boldsymbol{a})$ as the matrix that represents a linear operator; see, e.g., Axler 1997.) Lines are sent by f to curves, but they become straight lines under $\mathbf{D}f(\boldsymbol{a})$. See Figure 1.5.

◇ **Example 1.29** Let $f: \mathbb{R}^m \to \mathbb{R}^n$ be defined by the equation

$$f(\boldsymbol{x}) = \mathbf{A} \cdot \boldsymbol{x} + \boldsymbol{b},$$

where $\mathbf{A}$ is an $n \times m$ matrix, and $\boldsymbol{a} \in \mathbb{R}^n$. Then

$$\begin{aligned} f(\boldsymbol{a}+\boldsymbol{h}) = \mathbf{A}\cdot(\boldsymbol{a}+\boldsymbol{h})+\boldsymbol{b} &= \mathbf{A}\cdot\boldsymbol{a}+\boldsymbol{b}+\mathbf{A}\cdot\boldsymbol{h} \\ &= f(\boldsymbol{a})+\mathbf{A}\cdot\boldsymbol{h}. \end{aligned}$$

Hence, $R_{\boldsymbol{a}}(\boldsymbol{h}) = [f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a}) - \mathbf{A}\cdot\boldsymbol{h}]/\|\boldsymbol{h}\| = \mathbf{0}$; that is, $\mathbf{D}f(\boldsymbol{a}) = \mathbf{A}$.

We now show that the definition of derivative is stronger than directional derivative. In particular, we have:

■ **Theorem 1.30** *Let $A \subset \mathbb{R}^m$; let $f: A \to \mathbb{R}^n$. If f is differentiable at $\boldsymbol{a}$, then f is continuous at $\boldsymbol{a}$.*

Proof. Differentiablity at $\boldsymbol{a}$ implies that

$$\begin{aligned} \|f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a})\| &= \big\| \mathbf{D}f(\boldsymbol{a})\cdot\boldsymbol{h} + \|\boldsymbol{h}\|\, R_{\boldsymbol{a}}(\boldsymbol{h}) \big\| \\ &\leqslant \|\mathbf{D}f(\boldsymbol{a})\|\cdot\|\boldsymbol{h}\| + \|R_{\boldsymbol{a}}(\boldsymbol{h})\|\cdot\|\boldsymbol{h}\| \\ &\to 0, \end{aligned}$$

as $\boldsymbol{a}+\boldsymbol{h} \to \boldsymbol{a}$, where the inequality follows from the Triangle Inequality and Cauchy-Schwartz Inequality (Theorem 1.7). □

However, there is a nice connection between directional derivative and derivative.

■ **Theorem 1.31** *Let $A \subset \mathbb{R}^m$; let $f: A \to \mathbb{R}^n$. If f is differentiable at $\boldsymbol{a}$, then all the directional derivatives of f at $\boldsymbol{a}$ exist, and*

$$f'(\boldsymbol{a};\boldsymbol{u}) = \mathbf{D}f(\boldsymbol{a})\cdot\boldsymbol{u}.$$

Proof. Fix any $\boldsymbol{u} \in \mathbb{R}^m$ and take $\boldsymbol{h} = t\boldsymbol{u}$. Then

$$\begin{aligned} \frac{f(\boldsymbol{a}+t\boldsymbol{u}) - f(\boldsymbol{a})}{t} &= \frac{\mathbf{D}f(\boldsymbol{a})\cdot(t\boldsymbol{u}) + \|t\boldsymbol{u}\|\, R_{\boldsymbol{a}}(t\boldsymbol{u})}{t} \\ &= \mathbf{D}f(\boldsymbol{a})\cdot\boldsymbol{u} + \frac{|t|\cdot\|\boldsymbol{u}\|}{t} R_{\boldsymbol{a}}(t\boldsymbol{u}). \end{aligned}$$

The last term converges to zero as $t \to 0$, which proves that $f'(\boldsymbol{a};\boldsymbol{u}) = \mathbf{D}f(\boldsymbol{a})\cdot\boldsymbol{u}$. □

Exercise 1.32 Define $f: \mathbb{R}^2 \to \mathbb{R}$ by setting

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

Show that f is not differentiable at $(0,0)$.

Exercise 1.33 Let $f: \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = \sqrt{|xy|}$. Show that f is not differentiable at $(0,0)$.

1.3 Partial Derivatives and the Jacobian

We now introduce the notion of the "partial derivatives" of a real-valued function. Let $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_m)$ be the standard basis of $\mathbb{R}^m$, i.e.,

$$\begin{aligned} \boldsymbol{e}_1 &= (1, 0, 0, \dots, 0), \\ \boldsymbol{e}_2 &= (0, 1, 0, \dots, 0), \\ &\dots \\ \boldsymbol{e}_m &= (0, 0, \dots, 0, 1). \end{aligned}$$

Definition 1.34 (Partial Derivatives) Let $A \subset \mathbb{R}^m$; let $f : A \to \mathbb{R}$. We define the j^{th} *partial derivative of* f *at* $\boldsymbol{a}$ to be the directional derivative of f at $\boldsymbol{a}$ with respect to the vector $\boldsymbol{e}_j$, provided this derivative exists; and we denote it by $\mathbf{D}_j f(\boldsymbol{a})$. That is,

$$\mathbf{D}_j f(\boldsymbol{a}) = \lim_{t \to 0} \frac{f(\boldsymbol{a} + t\boldsymbol{e}_j) - f(\boldsymbol{a})}{t}.$$

Remark 1.35 It is important to note that $\mathbf{D}_j f(\boldsymbol{a})$ is the ordinary derivative of a certain function; in fact, if $g(x) = f(a_1, \dots, a_{j-1}, x, a_{j+1}, \dots, a_m)$, then $\mathbf{D}_j f(\boldsymbol{a}) = g'(a_j)$. This means that $\mathbf{D}_j f(\boldsymbol{a})$ is the slope of the tangent line at $(\boldsymbol{a}, f(\boldsymbol{a}))$ to the curve obtained by intersecting the graph of f with the plane $x_i = a_i$ with $i \neq j$. See Figure 1.6.

The following theorem relates partial derivatives to the derivative in the case where f is a real-valued function.

Theorem 1.36 *Let* $A \subset \mathbb{R}^m$*; let* $f : A \to \mathbb{R}$*. If* f *is differentiable at* $\boldsymbol{a}$*, then*

$$\mathbf{D} f(\boldsymbol{a}) = \begin{bmatrix} \mathbf{D}_1 f(\boldsymbol{a}) & \mathbf{D}_2 f(\boldsymbol{a}) & \cdots & \mathbf{D}_m f(\boldsymbol{a}) \end{bmatrix}.$$

Proof. If f is differentiable at $\boldsymbol{a}$, then $\mathbf{D} f(\boldsymbol{a})$ is a $(1 \times m)$-matrix. Let

$$\mathbf{D} f(\boldsymbol{a}) = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_m \end{bmatrix}.$$

It follows from Theorem 1.31 that

$$\mathbf{D}_j f(\boldsymbol{a}) = f'(\boldsymbol{a}; \boldsymbol{e}_j) = \mathbf{D} f(\boldsymbol{a}) \cdot \boldsymbol{e}_j = \lambda_j. \qquad \square$$

1.4 The Chain Rule

We extend the familiar chain rule to the current setting.

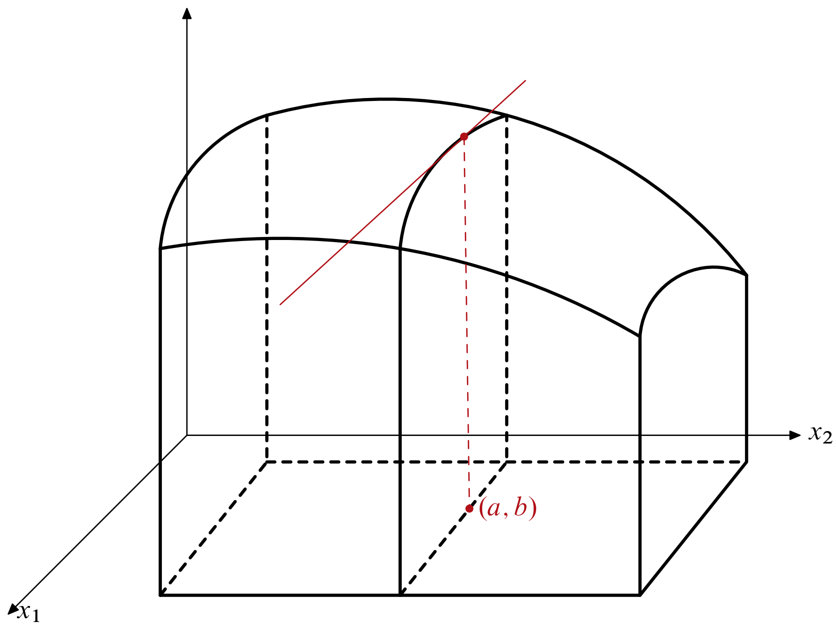


Figure 1.6: $\mathbf{D}_1 f(a, b)$.

■ **Theorem 1.37** (Chain Rule) *Let $A \subset \mathbb{R}^m$; let $B \subset \mathbb{R}^n$. Let $f : A \to \mathbb{R}^n$ and $g : B \to \mathbb{R}^p$, with $f(A) \subset B$. Suppose $f(\boldsymbol{a}) = \boldsymbol{b}$. If f is differentiable at $\boldsymbol{a}$, and if g is differentiable at $\boldsymbol{b}$, then the composite function $g \circ f : A \to \mathbb{R}^p$ is differentiable at $\boldsymbol{a}$. Furthermore,*

$$\mathbf{D}(g \circ f)(\boldsymbol{a}) = \mathbf{D}g(\boldsymbol{b}) \cdot \mathbf{D}f(\boldsymbol{a}).$$

Proof. Omitted. See, e.g., Spivak (1965, Theorem 2-2), Rudin (1976, Theorem 9.15), or Munkres (1991, Theorem 7.1). □

Here is an application of the Chain Rule.

■ **Theorem 1.38** *Let $A \subset \mathbb{R}^m$; let $f : A \to \mathbb{R}^n$. Suppose A contains a neighborhood of $\boldsymbol{a}$. Let $f_i : A \to \mathbb{R}$ be the i^{th} component function of f, so that*

$$f(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_n(\boldsymbol{x}) \end{bmatrix}.$$

(**a**) *The function f is differentiable at $\boldsymbol{a}$ if and only if each component function f_i is differentiable at $\boldsymbol{a}$.*

(b) *If f is differentiable at $\boldsymbol{a}$, then its derivative is the $(n \times m)$-matrix whose i^{th} row is the derivative of the function f_i. That is,*

$$\mathbf{D} f(\boldsymbol{a}) = \begin{bmatrix} \mathbf{D} f_1(\boldsymbol{a}) \\ \vdots \\ \mathbf{D} f_n(\boldsymbol{a}) \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 f_1(\boldsymbol{a}) & \cdots & \mathbf{D}_m f_1(\boldsymbol{a}) \\ \vdots & \ddots & \vdots \\ \mathbf{D}_1 f_n(\boldsymbol{a}) & \cdots & \mathbf{D}_m f_n(\boldsymbol{a}) \end{bmatrix}.$$

Proof. **(a)** Assume that f is differentiable at $\boldsymbol{a}$ and express the i^{th} component of f as

$$f_i = \pi_i \circ f,$$

where $\pi_i \colon \mathbb{R}^n \to \mathbb{R}$ is the projection that sends a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ to x_i. Notice that we can write π_i as $\pi_i(\boldsymbol{x}) = \mathbf{A} \cdot \boldsymbol{x}$, where $\mathbf{A}$ is a $1 \times n$ matrix such that

$$\mathbf{A} = \begin{bmatrix} 0 & \cdots 0 & 1 & 0 & \cdots & 0 \end{bmatrix},$$

where the number 1 appears at the i^{th} place. Then π_i is differentiable and $\mathbf{D}\pi(\boldsymbol{x}) = \mathbf{A}$ for all $\boldsymbol{x} \in A$ (see Example 1.29). By the Chain Rule, f_i is differentiable at $\boldsymbol{a}$ and

$$\mathbf{D} f_i(\boldsymbol{a}) = \mathbf{D}(\pi_i \circ f)(\boldsymbol{a}) = \mathbf{D}\pi_i\big(f(\boldsymbol{a})\big) \cdot \mathbf{D} f(\boldsymbol{a}) = \mathbf{A} \cdot \mathbf{D} f(\boldsymbol{a}). \tag{1.4}$$

Now suppose that each f_i is differentiable at $\boldsymbol{a}$. Let

$$\mathbf{B} := \begin{bmatrix} \mathbf{D} f_1(\boldsymbol{a}) \\ \vdots \\ \mathbf{D} f_n(\boldsymbol{a}) \end{bmatrix}.$$

We show that $\mathbf{D} f(\boldsymbol{a}) = \mathbf{B}$.

$$\begin{aligned} \|\boldsymbol{h}\|\, R_{\boldsymbol{a}}(\boldsymbol{h}) = f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a}) - \mathbf{B} \cdot \boldsymbol{h} &= \begin{bmatrix} f_1(\boldsymbol{a}+\boldsymbol{h}) - f_1(\boldsymbol{a}) - \mathbf{D} f_1 \cdot \boldsymbol{h} \\ \vdots \\ f_n(\boldsymbol{a}+\boldsymbol{h}) - f_n(\boldsymbol{a}) - \mathbf{D} f_n \cdot \boldsymbol{h} \end{bmatrix} \\ &= \|\boldsymbol{h}\| \begin{bmatrix} R_{\boldsymbol{a}}(\boldsymbol{h}; f_1) \\ \vdots \\ R_{\boldsymbol{a}}(\boldsymbol{h}; f_n), \end{bmatrix} \end{aligned}$$

where $R_{\boldsymbol{a}}(\boldsymbol{h}; f_i)$ is the Taylor remainder for f_i. It is clear that $\lim_{\boldsymbol{h}\to\mathbf{0}} R_{\boldsymbol{a}}(\boldsymbol{h}) = \mathbf{0}$, and which proves that $\mathbf{D} f(\boldsymbol{a}) = \mathbf{B}$.

(b) This claim follows from the previous part and Theorem 1.36. □

Remark 1.39 Theorem 1.38 implies that there is little loss of generality assuming $n = 1$, i.e., that our functions are real-valued. Multidimensionality of the *domain*, not the *range*, is what distinguished multivariable calculus from one-variable calculus.

▶ **Definition 1.40** (Jocobian Matrix) Let $A \subset \mathbb{R}^m$; let $f : A \to \mathbb{R}^n$. If the partial derivatives of the component functions f_i of f exist at $\boldsymbol{a}$, then one can form the matrix that has $\mathbf{D}_j f_i(\boldsymbol{a})$ as its entry in row i and column j. This matrix, denoted by $\mathbf{J} f(\boldsymbol{a})$, is called the *Jacobian matrix* of f. That is,

$$\mathbf{J} f(\boldsymbol{a}) = \begin{bmatrix} \mathbf{D}_1 f_1(\boldsymbol{a}) & \cdots & \mathbf{D}_m f_1(\boldsymbol{a}) \\ \vdots & \ddots & \vdots \\ \mathbf{D}_1 f_n(\boldsymbol{a}) & \cdots & \mathbf{D}_m f_n(\boldsymbol{a}) \end{bmatrix}.$$

Remark 1.41 **[1]** The Jacobian encapsulates all the essential information regarding the linear function that best approximates a differentiable function at a particular point. For this reason it is the Jacobian which is usually used in practical calculations with the derivative

[2] If f is differentiable at $\boldsymbol{a}$, then $\mathbf{J} f(\boldsymbol{a}) = \mathbf{D} f(\boldsymbol{a})$. However, it is possible for the partial derivatives, and hence the Jacobian matrix, to exist, without it following that f is differentiable at $\boldsymbol{a}$ (see Exercise 1.32).

1.5 The Implicit Function Theorem

Let $U \subset \mathbb{R}^k \times \mathbb{R}^n$ be open. Let $f : U \to \mathbb{R}^n$. Fix a point $(\boldsymbol{a}, \boldsymbol{b}) \in U$ and write $f(\boldsymbol{a}, \boldsymbol{b}) = \mathbf{0}$. Our goal is to solve the equation

$$f(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0} \tag{1.5}$$

near $(\boldsymbol{a}, \boldsymbol{b})$. More precisely, we hope to show that the set of points $(\boldsymbol{x}, \boldsymbol{y})$ nearby $(\boldsymbol{a}, \boldsymbol{b})$ at which $f(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0}$, the level-set of f through $\mathbf{0}$, is the graph of a function $\boldsymbol{y} = g(\boldsymbol{x})$. If so, g is the *implicit function* defined by (1.5). See Figure 1.7.

Under various hypotheses we will show that g exists, is unique, and is differentiable. Let us first consider an example.

◇ **Example 1.42** Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = x^2 + y^2 - 1.$$

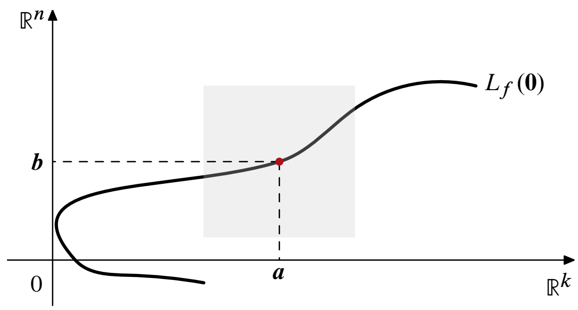


Figure 1.7: Near $(\boldsymbol{a}, \boldsymbol{b})$, $L_f(\mathbf{0})$ is the graph of a function $\boldsymbol{y} = g(\boldsymbol{x})$.

If we choose (a, b) with $f(a, b) = 0$ and $a \neq \pm 1$, there are (Figure 1.8) open intervals A containing a and B containing b with the following property: if $x \in A$, there is a unique $y \in B$ with $f(x, y) = 0$. We can therefore define a function $g\colon A \to \mathbb{R}$ by the condition $g(x) \in B$ and $f(x, g(x)) = 0$ (if $b > 0$, as indicated in Figure 1.8, then $g(x) = \sqrt{1 - x^2}$). For the function f we are considering there is another number b_1 such that $f(a, b_1) = 0$. There will also be an interval B_1 containing b_1 such that, when $x \in A$, we have $f(x, g_1(x)) = 0$ for a unique $g_1(x) \in B_1$ (here $g_1(x) = -\sqrt{1 - x^2}$). Both g and g_1 are differentiable. These functions are said to be defined *implicitly* by the equation $f(x, y) = 0$.

If we choose $a = 1$ or -1 it is impossible to find any such function g defined in an open interval containing a.

■ **Theorem 1.43** (Implicit Function Theorem) *Let $U \subset \mathbb{R}^{k+n}$ be open; let $f : U \to \mathbb{R}^n$ be of class $\mathscr{C}^r$. Write f in the form $f(\boldsymbol{x}, \boldsymbol{y})$ for $\boldsymbol{x} \in \mathbb{R}^k$ and $\boldsymbol{y} \in \mathbb{R}^n$. Suppose that $(\boldsymbol{a}, \boldsymbol{b})$ is a point of U such that $f(\boldsymbol{a}, \boldsymbol{b}) = \mathbf{0}$. Let $\mathbf{M}$ be the $n \times n$ matrix*

$$\mathbf{M} = \begin{bmatrix} \mathbf{D}_{k+1} f_1(\boldsymbol{a}, \boldsymbol{b}) & \mathbf{D}_{k+2} f_1(\boldsymbol{a}, \boldsymbol{b}) & \cdots & \mathbf{D}_{k+n} f_1(\boldsymbol{a}, \boldsymbol{b}) \\ \mathbf{D}_{k+1} f_2(\boldsymbol{a}, \boldsymbol{b}) & \mathbf{D}_{k+2} f_1(\boldsymbol{a}, \boldsymbol{b}) & \cdots & \mathbf{D}_{k+n} f_2(\boldsymbol{a}, \boldsymbol{b}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_{k+1} f_n(\boldsymbol{a}, \boldsymbol{b}) & \mathbf{D}_{k+2} f_n(\boldsymbol{a}, \boldsymbol{b}) & \cdots & \mathbf{D}_{k+n} f_n(\boldsymbol{a}, \boldsymbol{b}) \end{bmatrix}.$$

If $\det(\mathbf{M}) \neq 0$*, then near $(\boldsymbol{a}, \boldsymbol{b})$, $L_f(\mathbf{0})$ is the graph of a unique function $\boldsymbol{y} = g(\boldsymbol{x})$. Besides, g is $\mathscr{C}^r$.*

Proof. The proof is too long to give here. You can find it from, e.g., Spivak (1965, Theorem 2-12), Rudin (1976, Theorem 9.28), Munkres (1991, Theorem 2.9.2), or Pugh (2002, Theorem5.22). □

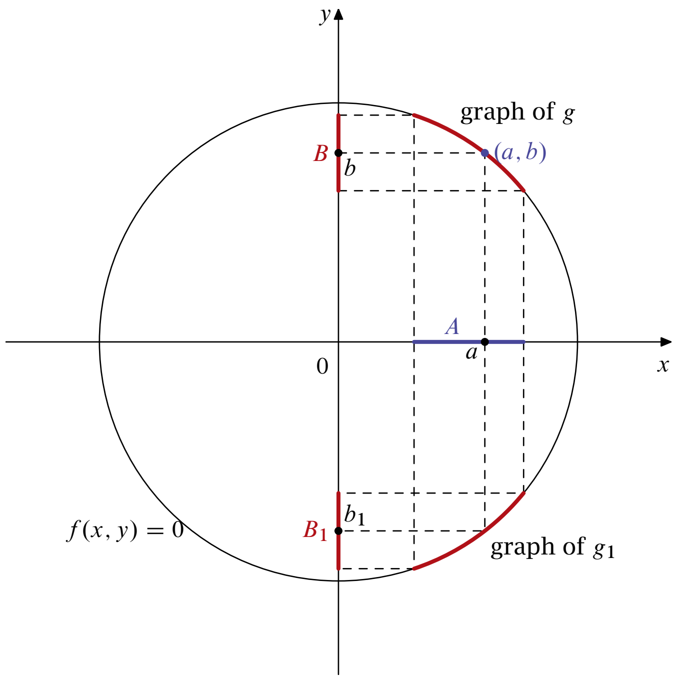


Figure 1.8: Implicit function theorem.

◇ **Example 1.44** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by the equation

$$f(x, y) = x^2 - y^3.$$

Then $(0, 0)$ is a solution of the equation $f(x, y) = 0$. Because $\partial f(0, 0)/\partial y = 0$, we do not expect to be able to solve this equation for y in terms of x near $(0, 0)$. But in fact, we can; and furthermore, the solution is unique! However, the function we obtain is not differentiable at $x = 0$. See Figure 1.9.

◇ **Example 1.45** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by the equation

$$f(x, y) = -x^4 + y^2.$$

Then $(0, 0)$ is a solution of the equation $f(x, y) = 0$. Because $\partial f(0, 0)/\partial y = 0$, we do not expect to be able to solve for y in terms of x near $(0, 0)$. In fact, however, we can do so, and we can do so in such a way that the resulting function is differentiable. However, the solution is not unique. See Figure 1.10.

Now the point $(1, 1)$ is also a solution to $f(x, y) = 0$. Because $\partial f(1, 1)/\partial y = 2$, one can solve this equation for y as a continuous function of x in a neighborhood of $x = 1$. See Figure 1.10.

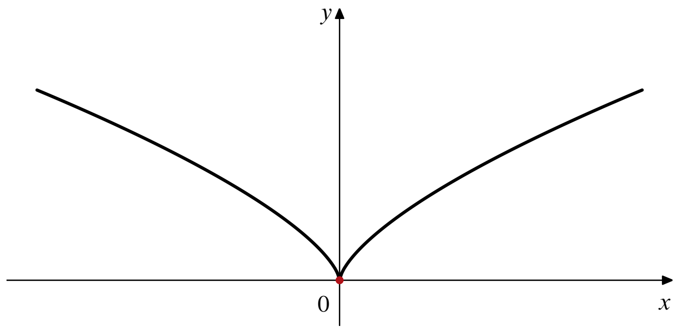


Figure 1.9: y is not differentiable at $x = 0$.

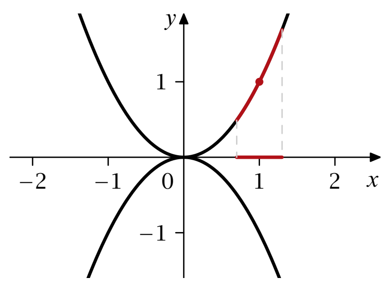


Figure 1.10: Example 1.45.

Remark 1.46 We will use the Implicit Function Theorem in Theorem 2.12. The theorem will also be used to derive comparative statics for economic models, which we perhaps do not have time to discuss.

1.6 Gradient and Its Properties

In this section we investigate the significance of the gradient vector, which is defined as follows:

▶ **Definition 1.47** (Gradient) Let $A \subset \mathbb{R}^m$; let $f : A \to \mathbb{R}$. Suppose A contains a neighborhood of $\boldsymbol{a}$. The *gradient* of f, denoted by $\nabla f(\boldsymbol{a})$, is defined by

$$\nabla f(\boldsymbol{a}) := \begin{bmatrix} \mathbf{D}_1 f(\boldsymbol{a}) & \mathbf{D}_2 f(\boldsymbol{a}) & \cdots & \mathbf{D}_m f(\boldsymbol{a}) \end{bmatrix}.$$

Remark 1.48 **[1]** It follows from Theorem 1.36 that if f is differentiable at $\boldsymbol{a}$, then $\nabla f(\boldsymbol{a}) = \mathbf{D} f(\boldsymbol{a})$. The inverse does not hold; see Remark 1.41[2].

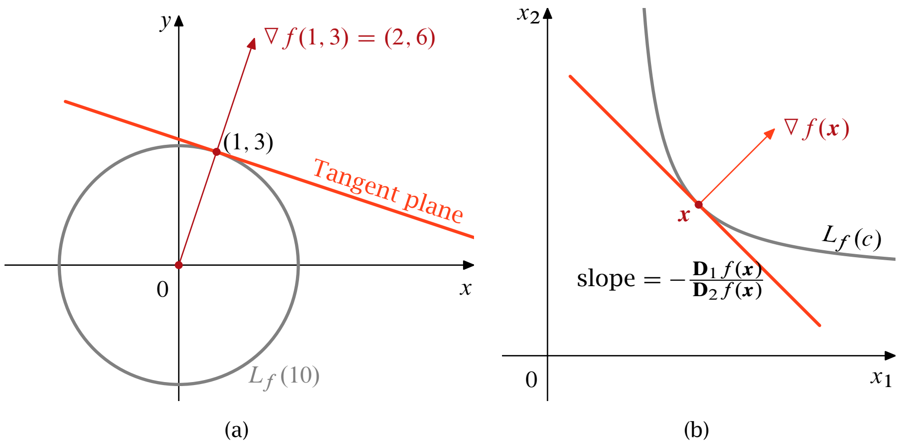


Figure 1.11: The geometric interpretation of $\nabla f(\boldsymbol{x})$.

[2] With the notation of gradient, we can write the Jocobian of $f = (f_1, \ldots, f_n)$ as

$$\mathbf{J} f(\boldsymbol{a}) = \begin{bmatrix} \nabla f_1(\boldsymbol{a}) \\ \vdots \\ \nabla f_n(\boldsymbol{a}) \end{bmatrix}.$$

Let $f: \mathbb{R}^m \to \mathbb{R}^n$ with $f = (f_1, \ldots, f_n)$. Recall that the *level set* of f through $\mathbf{0}$ is given by

$$L_f(\mathbf{0}) := \left\{\boldsymbol{x} \in \mathbb{R}^m : f(\boldsymbol{x}) = \mathbf{0}\right\} = \bigcap_{i=1}^{n} L_{f_i}(0). \tag{1.6}$$

Given a point $\boldsymbol{a} \in L_f(\mathbf{0})$, it is intuitively clear what it means for a plane to be tangent to $L_f(\mathbf{0})$ at $\boldsymbol{a}$. Figure 1.12 shows some examples of tangent planes. A formal definition of tangent plane will be given later. In this section we show that the gradient $\nabla f(\boldsymbol{a})$ is orthogonal to the tangent plane of $L_f(\mathbf{0})$ at $\boldsymbol{a}$ and, under some conditions, the tangent plane of $L_f(\mathbf{0})$ at $\boldsymbol{a}$ can be characterized by the vectors that are orthogonal to $\nabla f(\boldsymbol{a})$. We begin with the simplest case that $f: \mathbb{R}^2 \to \mathbb{R}$. Here is an example.

◇ **Example 1.49** Let $f(x, y) = x^2 + y^2$. The level set $L_f(10)$ is given by

$$L_f(10) = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 10\right\}.$$

Calculus yields

$$\left.\frac{\mathrm{d}y}{\mathrm{d}x}\right|_{\text{along } L_f(10) \text{ at } (1,3)} = -\frac{1}{3}.$$

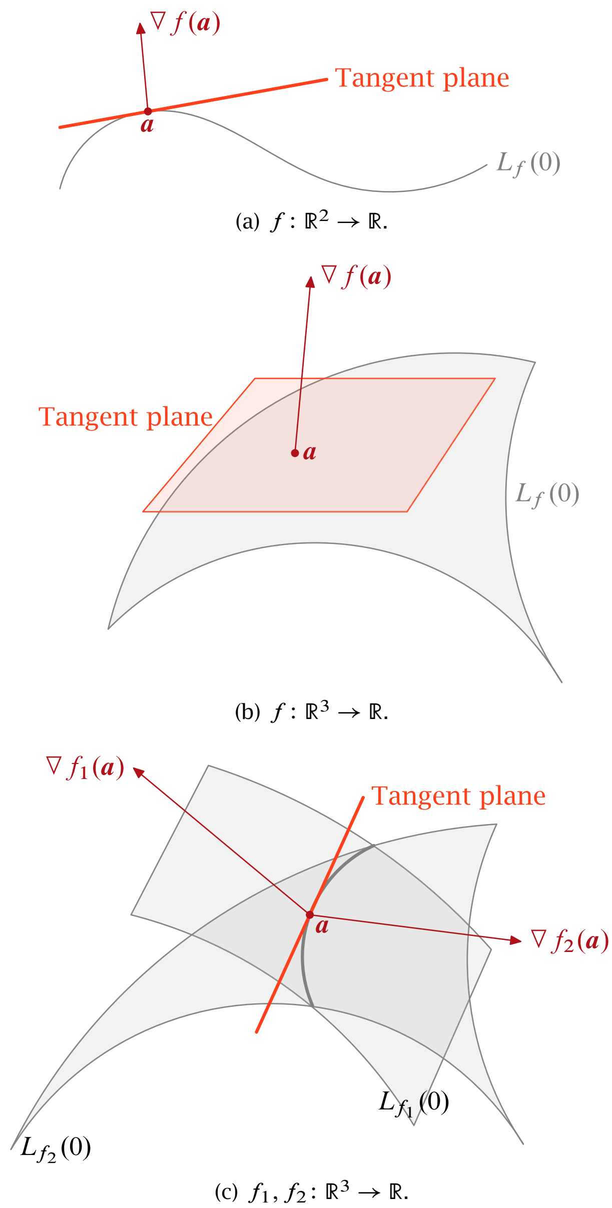


Figure 1.12: Examples of tangent planes.

Hence, the tangent plane at $(1,3)$ is given by $y = 3 - (x-1)/3$. Since $\nabla f(1,3) = (2,6)$, the result follows immediately; see Figure 1.11(a).

The result in Example 1.49 can be explained as follows. If we change x_1 and x_2, and are to remain on $L_f(0)$, then $\mathrm{d}x_1$ and $\mathrm{d}x_2$ must be such as to leave the value of f unchanged at 0. They must therefore satisfy

$$f'(\boldsymbol{x};(\mathrm{d}x_1, \mathrm{d}x_2)) = \mathbf{D}_1 f(\boldsymbol{x})\,\mathrm{d}x_1 + \mathbf{D}_2 f(\boldsymbol{x})\,\mathrm{d}x_2 = 0. \tag{1.7}$$

By solving (1.7) for $\mathrm{d}x_2/\mathrm{d}x_1$, the slope of the level set through $\boldsymbol{x}$ will be (see Figure 1.11(b))

$$\frac{\mathrm{d}x_2}{\mathrm{d}x_1} = -\frac{\mathbf{D}_1 f(\boldsymbol{x})}{\mathbf{D}_2 f(\boldsymbol{x})}.$$

Since the slope of the vector $\nabla f(\boldsymbol{x}) = (\mathbf{D}_1 f(\boldsymbol{x}), \mathbf{D}_2 f(\boldsymbol{x}))$ is $\mathbf{D}_2 f(\boldsymbol{x})/\mathbf{D}_1 f(\boldsymbol{x})$, we obtain the desired result.

We then present the general result. For simplicity, we assume throughout this section that each $f_i \in \mathcal{C}^1$. For $L_f(\mathbf{0})$ defined in (1.6), obtaining an explicit representation for the tangent plane is a fundamental problem that we now address. First we define *curves* on $L_f(\mathbf{0})$ and the *tangent plane* at some point $\boldsymbol{x} \in \mathbb{R}^m$. You may want to refer some Differential Geometry textbooks, e.g., O'neill (2006), Spivak (1999), or Lee (2009), for understanding some of the following concepts better.

One can picture a *curve* in $\mathbb{R}^m$ as a trip taken by a moving point $\boldsymbol{c}$. At each "time" t in some interval $[a,b] \subset \mathbb{R}$, $\boldsymbol{c}$ is located at the point

$$\boldsymbol{c}(t) = \big(c_1(t), \ldots, c_m(t)\big) \in \mathbb{R}^m.$$

In rigorous terms then, $\boldsymbol{c}$ is a function from $[a,b]$ to $\mathbb{R}^m$, and the component functions $c_1, \ldots, c_m$ are its *Euclidean coordinate functions*. We define the function $\boldsymbol{c}$ to be differentiable provided its component functions are differentiable.

◇ **Example 1.50** A *helix* $\boldsymbol{c}\colon \mathbb{R} \to \mathbb{R}^3$ is obtained through the formula

$$\boldsymbol{c}(t) = (a\cos t, a\sin t, bt),$$

where $a, b > 0$. See Figure 1.13.

▶ **Definition 1.51** A *curve on* $L_f(\mathbf{0})$ is a continuous curve $\boldsymbol{c}\colon [a,b] \to L_f(\mathbf{0})$. A curve $\boldsymbol{c}(t)$ is said to pass through the point $\boldsymbol{a} \in L_f(\mathbf{0})$ if $\boldsymbol{a} = \boldsymbol{c}(t^*)$ for some $t^* \in [a,b]$.

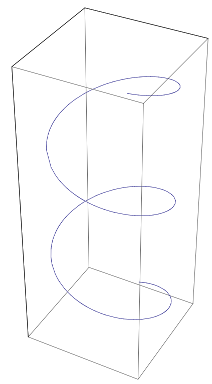


Figure 1.13: The Helix.

▶ **Definition 1.52** The *tangent plane* at $\boldsymbol{a} \in L_f(\mathbf{0})$, denoted $T_f(\boldsymbol{a})$, is defined as the collection of the derivatives at $\boldsymbol{a}$ of all differentiable curves on $L_f(\mathbf{0})$ passing through $\boldsymbol{a}$.

Ideally, we would like to express the tangent plane defined in Definition 1.52 in terms of derivatives of functions f_i that defines the surface $L_f(\mathbf{0})$ (see Example 1.49). We introduce the subspace

$$M := \left\{\boldsymbol{x} \in \mathbb{R}^m : \mathbf{D} f(\boldsymbol{a}) \cdot \boldsymbol{x} = \mathbf{0}\right\} \tag{1.8}$$

and investigate under what conditions M is equal to the tangent plane at $\boldsymbol{a}$. The following result shows that $\nabla f_i(\boldsymbol{a})$ is orthogonal to the tangent plane $T_f(\boldsymbol{a})$ for all $\boldsymbol{a} \in L_f(\mathbf{0})$.

■ **Theorem 1.53** *For each $\boldsymbol{a} \in L_f(\mathbf{0})$, the gradient $\nabla f_i(\boldsymbol{a})$ is orthogonal to the tangent plane $T_f(\boldsymbol{a})$.*

Proof. We establish this result by showing $T_f(\boldsymbol{a}) \subset M$ for each $\boldsymbol{a} \in L_f(\mathbf{0})$. Every curve $\boldsymbol{c}(t)$ passing through $\boldsymbol{a}$ at $t = t^*$ satisfies $f(\boldsymbol{x}(t*)) = \mathbf{0}$, and so

$$\mathbf{D} f\left(\boldsymbol{c}(t^*)\right) \cdot \mathbf{D}\boldsymbol{c}(t^*) = \mathbf{0}.$$

That is, $\mathbf{D}\boldsymbol{c}(t^*) \in M$. □

▶ **Definition 1.54** A point $\boldsymbol{a} \in L_f(\mathbf{0})$ is said to be a *regular point* if the gradient vectors $(\nabla f_1(\boldsymbol{a}), \ldots, \nabla f_n(\boldsymbol{a}))$ are linearly independent.

In general, at regular points it is possible to characterize the tangent plane in terms of $\nabla f_1(\boldsymbol{a}), \ldots, \nabla f_n(\boldsymbol{a})$.

■ **Theorem 1.55** *At a regular point $\boldsymbol{a}$ of $L_f(\mathbf{0})$, the tangent plane $T_f(\boldsymbol{a})$ is equal to M*

Proof. We show that $M \subset T_f(\boldsymbol{a})$. Combining this result with Theorem 1.53, we have $T_f(\boldsymbol{a}) = M$.

To show $M \subset T_f(\boldsymbol{a})$, we must show that if $\boldsymbol{x} \in M$ then there exists a curve on $L_f(\boldsymbol{a})$ passing through $\boldsymbol{a}$ with derivative $\boldsymbol{x}$. To construct such a curve we consider the equations

$$f\left(\boldsymbol{a} + t\boldsymbol{x} + \mathbf{D}f(\boldsymbol{a})^T \cdot u(t)\right) = \mathbf{0}, \tag{1.9}$$

where for fixed t we consider $u(t) \in \mathbb{R}^n$ to be the unknown. This is a nonlinear system of n equations and n unknowns, parametrized continuously by t. At $t = 0$ there is a solution $u(0) = \mathbf{0}$. The Jacobian matrix of the system with respect to u at $t = 0$ is the $n \times n$ matrix

$$\mathbf{D}f(\boldsymbol{a}) \cdot \mathbf{D}f(\boldsymbol{a})^T,$$

which is nonsingular, since $\mathbf{D}f(\boldsymbol{a})$ is of full rank if $\boldsymbol{a}$ is a regular point. Thus, by the Implicit Function Theorem (Theorem 1.43) there is a continuously differentiable solution $u(t)$ in some region $t \in [-a, a]$.

The curve

$$\boldsymbol{c}(t) := \boldsymbol{a} + t\boldsymbol{x} + \mathbf{D}f(\boldsymbol{a})^T \cdot u(t)$$

is thus a curve on $L_f(\mathbf{0})$. By differentiating the system (1.9) with respect to t at $t = 0$ we obtain

$$\mathbf{D}f(\boldsymbol{a}) \cdot \left[\boldsymbol{x} + \mathbf{D}f(\boldsymbol{a})^T \cdot \mathbf{D}u(0)\right] = \mathbf{0}. \tag{1.10}$$

By definition of $\boldsymbol{x}$ we have $\mathbf{D}f(\boldsymbol{a})\cdot\boldsymbol{x} = \mathbf{0}$ and thus, again since $\mathbf{D}f(\boldsymbol{a})\cdot\mathbf{D}f(\boldsymbol{a})^T$ is nonsingular, we conclude from (1.10) that

$$\mathbf{D}u(0) = \left[\mathbf{D}f(\boldsymbol{a}) \cdot \mathbf{D}f(\boldsymbol{a})^T\right]^{-1} \cdot \mathbf{0} = \mathbf{0}.$$

Therefore,

$$\mathbf{D}\boldsymbol{c}(0) = \boldsymbol{x} + \mathbf{D}f(a)^T \cdot \mathbf{D}u(0) = \boldsymbol{x},$$

and the constructed curve has derivative $\boldsymbol{x}$ at $\boldsymbol{a}$. □

◇ **Example 1.56** **[1]** In $\mathbb{R}^2$ let $f(x_1, x_2) = x_1$. Then $L_f(0)$ is the x_2-axis, and every point on that asis is regular since $\nabla f(0, x_2) = (1, 0)$. In this case, $T_f((x_1, 0)) = M$, and which is the x_2-axis.

[2] In $\mathbb{R}^2$ let $f(x_1, x_2) = x_1^2$. Again, $L_f(0)$ is the x_2-axis, but now no point on the x_2-axis is regular: $\nabla f(0, x_2) = (0, 0)$. Indeed in this case $M = \mathbb{R}^2$, while the tangent plane is the x_2-axis.

We close this section by providing another property of gradient vectors. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function and $\boldsymbol{a} \in \mathbb{R}^n$, where $\nabla f(\boldsymbol{a}) \neq \mathbf{0}$. Suppose that we want to determine the direction in which f increases most rapidly at $\boldsymbol{a}$. By a "direction" here we mean a unit vector $\boldsymbol{u}$. Let $\theta_{\boldsymbol{u}}$ denote the angle between $\boldsymbol{u}$ and $\nabla f(\boldsymbol{a})$. Then

$$f'(\boldsymbol{a}; \boldsymbol{u}) = \nabla f(\boldsymbol{a}) \cdot \boldsymbol{u} = \|\nabla f(\boldsymbol{a})\| \cos \theta_{\boldsymbol{u}}.$$

But $\cos \theta_{\boldsymbol{u}}$ attains its maximum value of 1 when $\theta_{\boldsymbol{u}} = 0$, that is, when $\boldsymbol{u}$ and $\nabla f(\boldsymbol{a})$ are collinear and point in the same direction. We conclude that $\|\nabla f(\boldsymbol{a})\|$ is the maximum value of $f'(\boldsymbol{a}; \boldsymbol{u})$ for $\boldsymbol{u}$ a unit vector, and that this maximum value is attained with $\boldsymbol{u} = \nabla f(\boldsymbol{a}) / \|\nabla f(\boldsymbol{a})\|$.

1.7 Continuously Differentiable Functions

We know that mere existence of the partial derivatives does not imply differentiability (see Exercise 1.32). If, however, we impose the additional condition that these partial derivatives are continuous, then differentiability is assured.

■ **Theorem 1.57** *Let A be open in $\mathbb{R}^m$. Suppose that the partial derivatives $\mathbf{D}_j f_i(\boldsymbol{x})$ of the component functions of f exist at each point $\boldsymbol{x} \in A$ and are continuous on A. Then f is differentiable at each point of A.*

A function satisfying the hypotheses of this theorem is often said to be *continuously differentiable*, or *of class $\mathcal{C}^1$*, on A.

Proof of Theorem 1.57. It suffices to show that each component function of f is differentiable by Theorem 1.38. Therefore we may restrict ourselves to the case of a real-valued function $f : A \to \mathbb{R}$. Let $\boldsymbol{a} \in A$. We claim that $\mathbf{D} f(\boldsymbol{a}) = \nabla f(\boldsymbol{a})$ when $f \in \mathcal{C}^1$.

Recall that $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ is the standard basis of $\mathbb{R}^m$. Then every $\boldsymbol{h} = (h_1, \ldots, h_m) \in \mathbb{R}^m$ can be written as $\boldsymbol{h} = \sum_{i=1}^m h_i \boldsymbol{e}_i$. For each $i = 1, \ldots, m$, let

$$\boldsymbol{p}_i := \boldsymbol{a} + \sum_{k=1}^{i} h_k \boldsymbol{e}_k = \boldsymbol{p}_{i-1} + h_i \boldsymbol{e}_i,$$

where $\boldsymbol{p}_0 := \boldsymbol{a}$. Figure 1.14 illustrates the case where $m = 3$ and all h_i are positive. For each $i = 1, \ldots, m$, we also define a function $\sigma_i : [0, 1] \to \mathbb{R}^m$ by

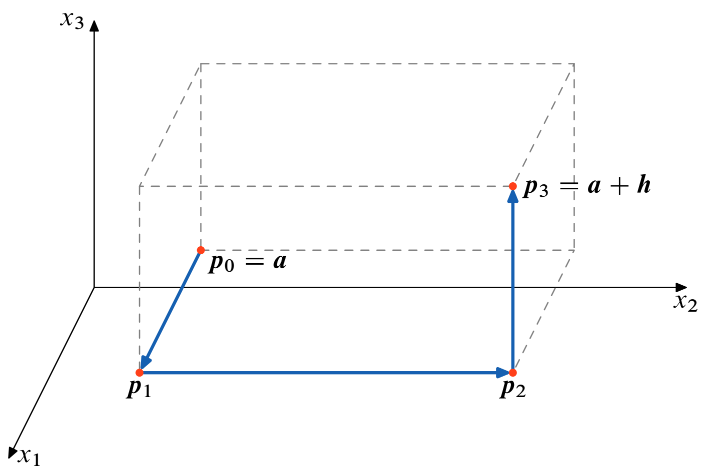


Figure 1.14: The segmented path from $\boldsymbol{a}$ to $\boldsymbol{a}+\boldsymbol{h}$.

letting

$$\sigma_i(t) = \boldsymbol{p}_{i-1} + t h_i \boldsymbol{e}_i.$$

So, σ_i is a segment from $\boldsymbol{p}_{i-1}$ to $\boldsymbol{p}_i$.

By the one-dimensional chain rule and mean value theorem applied to the differentiable real-valued function $g\colon [0,1] \to \mathbb{R}$ defined by

$$g(t) = (f \circ \sigma_i)(t),$$

there exists $\bar{t}_i \in (0,1)$ such that

$$\begin{aligned}
f(\boldsymbol{p}_i) - f(\boldsymbol{p}_{i-1}) &= g(1) - g(0) \\
&= g'\left(\bar{t}_i\right) \\
&= \left.\frac{\mathrm{d} f\left(h_1^{i-1}, h_{i-1}^{i-1}, h_i^{i-1} + t h_i, h_{i+1}^{i-1}, \ldots, h_m^{i-1}\right)}{\mathrm{d}t}\right|_{t=\bar{t}_i} \\
&= \mathbf{D}_i f\left(\sigma_i\left(\bar{t}_i\right)\right) \cdot h_i.
\end{aligned}$$

Telescoping $f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a})$ along $(\sigma_1, \ldots, \sigma_m)$ gives

$$\begin{aligned}
f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a}) - \nabla f(\boldsymbol{a}) \cdot \boldsymbol{h} &= \sum_{i=1}^{m} \left[f\left(\boldsymbol{p}_i\right) - f\left(\boldsymbol{p}_{i-1}\right) \right] - \nabla f(\boldsymbol{a}) \cdot \boldsymbol{h} \\
&= \sum_{i=1}^{m} \left[\mathbf{D}_i f\left(\sigma_i\left(\bar{t}_i\right)\right) - \mathbf{D}_i f(\boldsymbol{a}) \right] \cdot h_i.
\end{aligned}$$

Continuity of the partials implies that $\mathbf{D}_i f(\sigma_i(\bar{t}_i)) - \mathbf{D}_i f(\boldsymbol{a}) \to 0$ as $\|\boldsymbol{h}\| \to 0$. $\square$

Remark 1.58 It follows from Theorem 1.57 that $\sin(xy)$ and $xy^2 + ze^{xy}$ are both differentiable since they are of class $\mathcal{C}^1$.

Let $A \subset \mathbb{R}^m$ and $f\colon A \to \mathbb{R}^n$. Suppose that the partial derivative $\mathbf{D}_j f_i$ of the component functions of f exist on A. These then are functions from A to $\mathbb{R}$, and we may consider their partial derivatives, which have the form

$$\mathbf{D}_k(\mathbf{D}_j f_i) =: \mathbf{D}_{jk} f_i$$

and are called the *second-order partial derivatives of* f. Similarly, one defines the third-order partial derivatives of the functions f_i, or more generally the *partial derivatives of order* r for arbitrary r.

▶ **Definition 1.59** If the partial derivatives of the function f_i of order less than or equal to r are continuous on A, we say f is *of class* $\mathcal{C}^r$ on A. We say f is *of class* $\mathcal{C}^\infty$ on A if the partials of the functions f_i of all orders are continuous on A.

▶ **Definition 1.60** (Hessian) Let $\boldsymbol{a} \in A \subset \mathbb{R}^m$; let $f\colon A \to \mathbb{R}$ be twice-differentiable at $\boldsymbol{a}$. The $m \times m$ matrix representing the second derivative of f is called the *Hessian* of f, denoted $\mathbf{H}f(\boldsymbol{a})$:

$$\mathbf{H}f(\boldsymbol{a}) = \begin{bmatrix} \mathbf{D}_{11}f(\boldsymbol{a}) & \mathbf{D}_{12}f(\boldsymbol{a}) & \cdots & \mathbf{D}_{1m}f(\boldsymbol{a}) \\ \mathbf{D}_{21}f(\boldsymbol{a}) & \mathbf{D}_{22}f(\boldsymbol{a}) & \cdots & \mathbf{D}_{2m}f(\boldsymbol{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{D}_{m1}f(\boldsymbol{a}) & \mathbf{D}_{m2}f(\boldsymbol{a}) & \cdots & \mathbf{D}_{mm}f(\boldsymbol{a}) \end{bmatrix} = \mathbf{D}(\nabla f).$$

Remark 1.61 If $f\colon A \to \mathbb{R}$ is of class $\mathcal{C}^2$, then the Hessian of f is a symmetric matrix, i.e., $\mathbf{D}_{ij}f(\boldsymbol{a}) = \mathbf{D}_{ji}f(\boldsymbol{a})$ for all $i, j = 1, \ldots, m$ and for all $\boldsymbol{a} \in A$. See Rudin (1976, Corollary to Theorem 9.41, p. 236).

Exercise 1.62 Find the Hessian of the Cobb-Douglas function

$$f(x, y) = x^\alpha y^\beta.$$

1.8 Quadratic Forms: Definite and Semidefinite Matrices

▶ **Definition 1.63** (Quadratic Form) Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. A *quadratic form* on $\mathbb{R}^n$ is a function $Q_{\mathbf{A}}\colon \mathbb{R}^n \to \mathbb{R}$ of the form

$$Q_{\mathbf{A}}(\boldsymbol{x}) = \boldsymbol{x} \cdot \mathbf{A}\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j.$$

Since the quadratic form $Q_{\mathbf{A}}$ is completely specified by the matrix $\mathbf{A}$, we henceforth refer to $\mathbf{A}$ itself as the quadratic form. Observe that if f is of class $\mathcal{C}^2$, then the Hessian $\mathbf{H}f$ of f defines a quadratic form; see Remark 1.61.

▶ **Definition 1.64** A quadratic form $\mathbf{A}$ is said to be

- *positive definite* if we have $\boldsymbol{x} \cdot \mathbf{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \in \mathbb{R}^n \smallsetminus \{\mathbf{0}\}$;
- *positive semidefinite* if we have $\boldsymbol{x} \cdot \mathbf{A}\boldsymbol{x} \geqslant 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$;
- *negative definite* if we have $\boldsymbol{x} \cdot \mathbf{A}\boldsymbol{x} < 0$ for all $\boldsymbol{x} \in \mathbb{R}^n \smallsetminus \{\mathbf{0}\}$;
- *negative semidefinite* if we have $\boldsymbol{x} \cdot \mathbf{A}\boldsymbol{x} \leqslant 0$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

1.9 Homogeneous Functions and Euler's Formula

▶ **Definition 1.65** (Homogeneous Function) A function $f : \mathbb{R}^n \to \mathbb{R}$ is *homogeneous of degree* r (for $r = \dots, -1, 0, 1, \dots$) if for every $t > 0$ we have

$$f(tx_1, \dots, tx_n) = t^r f(x_1, \dots, x_n).$$

? **Exercise 1.66** The function

$$f(x, y) = Ax^\alpha y^\beta, \quad A, \alpha, \beta > 0,$$

is known as the *Cobb-Douglas* function. Check whether this function is homogeneous.

■ **Theorem 1.67** (Euler's Formula) *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is homogeneous of degree r (for some $r = \dots, -1, 0, 1, \dots$) and differentiable. Then at any $\boldsymbol{x}^* \in \mathbb{R}^n$ we have*

$$\nabla f(\boldsymbol{x}^*) \cdot \boldsymbol{x}^* = rf(\boldsymbol{x}^*).$$

Proof. By definition we have

$$f(t\boldsymbol{x}^*) - t^r f(\boldsymbol{x}^*) = 0.$$

Differentiating with respect to t using the chain rule, we have

$$\nabla f(t\boldsymbol{x}^*) \cdot \boldsymbol{x}^* = rt^{r-1} f(\boldsymbol{x}^*).$$

Evaluating at $t = 1$ gives the desired result. □

■ **Lemma 1.68** *If f is homogeneous of degree r, its partial derivatives are homogeneous of degree $r - 1$.*