Chapter 5

Differentiability of Functions

Partial differentiation and differentiation of functions of several variables are discussed in Sections 1 and 2 respectively. Relations among continuity, existence of partial derivatives, and differentiability are crucial for all subsequent developments. The Chain Rule as the main technical tool in differentiation is taken up in Section 3. The gradient vector and directional derivatives are introduced in Section 4. This chapter ends with some applications of the differential of a function in Section 5.

5.1 Partial Differentiation

We first recall how the derivative of a function of a single variable was defined. Let f be a function defined on some interval (a, b) and $x_0 \in (a, b)$. In calculus the derivative of f at x_0 is defined to be the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} .$$

Whenever the limit exists, f is said to have a derivative at x_0 and this limit is denoted by $f'(x_0)$ or $df/dx(x_0)$. A function is continuous at x_0 whenever its derivative exists at this point. In other words, differentiability is a property stronger than continuity. In fact, the absolute value function $x \mapsto |x|$ is continuous but does not have a derivative at 0, showing that these two properties are not equivalent.

Now, we consider the situation in higher dimension. Let G be an open set in $\mathbb{R}^n$ and f a function defined in G. For $z \in G$, the function $\varphi(t) = f(t, z_2, \cdots, z_n),\ z = (z_1, \cdots, z_n)$, is a function defined on some open interval containing z_1 and $\varphi(z_1) = f(z)$. The function f is said to have a **partial derivative** in x_1 if $\varphi(t)$ has a derivative at z_1. In notation we have

$$\begin{aligned}\frac{\partial f}{\partial x_1}(z) &= \lim_{t\to z_1}\frac{\varphi(t)-\varphi(z_1)}{t-z_1}\\ &= \lim_{x_1\to z_1}\frac{f(x_1,z_2,\cdots,z_n)-f(z_1,z_2,\cdots,z_n)}{x_1-z_1}\ .\end{aligned}$$

Similarly, one can define the partial derivative in x_j for $j=2,\cdots,n$. We say a function has partial derivatives in G if the partial derivatives $\partial f/\partial x_j, j=1,\cdots,n$, exist at every point of G.

In calculus we learned that the derivative of a function is the instantaneous rate of change of the function. For a multi-variable function $f(x_1,\cdots,x_n)$ the partial derivative in x_j is the instantaneous rate of change of f in the j-th coordinate while the other coordinates are unchanged. Here are some examples:

- Let $H(x,y)$ be the height of the mountain from sea level. In other words, the mountain is given by the graph $\{(x,y,H(x,y))\}$. The partial derivative $\partial H/\partial x(x,y)$ is the instantaneous rate of change when one moves parallel to the x-axis. The height $H(x_0+h,y_0)$ would be approximately equal to $H(x_0,y_0)+H_x(x_0,y_0)h$ with improving accuracy as h shrinks.

- Let $u(x,y,z)$ be the spatial temperature distribution at a fixed time. Then u_x at (x_0,y_0,z_0) is the instantaneous rate of change of the temperature, which means that the temperature at (x_1,y_0,z_0) is approximately given by

  $$u(x_0,y_0,z_0)+\frac{\partial u}{\partial x}(x_0,y_0,z_0)(x_1-x_0)$$

  when x_1-x_0 is small.

- Let $P(x,y)$ be the profit incurred from selling x units of Product A and y units of Product B. Then $P_x(x,y)$ is a rough measure on the increases or decreases in the profit when $x+1$ units of Product A are sold while the units sold in Product B are fixed.

- In economics the productivity is a function $F(x,y)$ where x is the units of labor and y is the units of capital. Then $F_x(x,y)$, the marginal productivity of labor, represents the instantaneous rate of change of productivity with respect to labor. Similarly, $F_y(x,y)$ is called the marginal productivity of capital.

From the definition you can see that partial derivatives are essentially the same as the derivative in the single variable case. When performing the partial derivative in x_j, you pretend all other variables x_k's, $k\neq j$, are fixed as constants. Let us look at some examples.

Example 5.1. Find the partial derivatives of

(a)

$$f(x,y) = \frac{xy^3}{x-y} \ ,$$

and

(b)

$$g(x,y,z) = e^{x/y} \sin xz \ .$$

We have

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

On the other hand,

$$\frac{\partial g}{\partial x} = \frac{e^{x/y} \sin xz}{y} + ze^{x/y} \cos xz \ ,$$
$$\frac{\partial g}{\partial y} = -\frac{e^{x/y} \sin xz}{y^2} \ ,$$
$$\frac{\partial g}{\partial z} = xe^{x/y} \cos xz \ .$$

Whenever the partial derivatives exist in G, they are again functions in G and one can consider their partial derivatives. In this way we obtain partial derivatives of higher order. For instance, we have

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) \ .$$
$$\frac{\partial^3 g}{\partial x_k \partial x_j \partial x_i} = \frac{\partial}{\partial x_k}\left(\frac{\partial}{\partial x_j}\left(\frac{\partial g}{\partial x_i}\right)\right) \ ,$$
$$\frac{\partial^4 h}{\partial x^2 \partial y \partial z} = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial h}{\partial z}\right)\right)\right) \ ,$$

and so on. In dimensions $n = 2, 3$, we often use subscripts to denote partial derivatives, for instance,

$$f_{xy} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} \ ,$$
$$g_{xyz} = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial y}\left(\frac{\partial g}{\partial x}\right)\right) = \frac{\partial^3 g}{\partial z \partial y \partial x} \ .$$

Be careful about the difference in the order of partial differentiation in these two notations.

Example 5.2. Verify that the function

$$f(x,t) = \frac{1}{2\pi t}\exp\left(-\frac{x^2+y^2}{4t}\right)$$

satisfies the heat equation

$$\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \ ,$$

We have

$$\frac{\partial f}{\partial t} = \frac{1}{2\pi}\left(\frac{-1}{t^2} + \frac{x^2+y^2}{4t^2}\right)\exp\left(-\frac{x^2+y^2}{4t}\right) \ ,$$

$$\frac{\partial f}{\partial x} = \frac{1}{2\pi}\frac{-x}{2t^2}\exp\left(-\frac{x^2+y^2}{4t}\right) \ ,$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{2\pi}\left(\frac{-1}{2t^2} + \frac{x^2}{4t^2}\right)\exp\left(-\frac{x^2+y^2}{4t}\right) \ .$$

Similarly,

$$\frac{\partial^2 f}{\partial y^2} = \frac{1}{2\pi}\left(\frac{-1}{2t^2} + \frac{y^2}{4t^2}\right)\exp\left(-\frac{x^2+y^2}{4t}\right) \ .$$

It follows that

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{1}{2\pi}\left(\frac{-1}{t^2} + \frac{x^2+y^2}{4t^2}\right)\exp\left(-\frac{x^2+y^2}{4t}\right) \ .$$

By comparing the expressions for f_t and $f_{xx} + f_{yy}$ we see that f satisfies the two dimensional heat equation.

Example 5.3. Let

$$\varphi(x,y) = y^2e^{2x} + 6x^2 - 2y \ .$$

We have

$$\varphi_x(x,y) = 2y^2e^{2x} + 12x \ , \quad \varphi_y(x,y) = 2ye^{2x} - 2 \ ,$$

and

$$\varphi_{xy}(x,y) = 4ye^{2x} \ , \quad \varphi_{yx}(x,y) = 4ye^{2x} \ .$$

Therefore, φ_{xy} and φ_{yx} are the same. Is it always true?

The answer is no. An example in which f_{xy} is not equal to f_{yx} can be found in the exercise. Nevertheless, the following result asserts that partial derivatives are independent of their order of taking under a mild continuity assumption. It saves us a lot of trouble.

Theorem 5.1. *Assume that f has partial derivatives up to second order in the open set G. Then for $z \in G$,*

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

provided these two partial derivatives are continuous at z.

In the example above φ_{xy} and φ_{yx} are clearly continuous. By this theorem they must be the same.

Proof. * Clearly we could take $(i, j) = (1, 2)$. In finding the partial derivatives only the first two components would change while the others remain constant. Therefore, it suffices to prove the proposition for $n = 2$. By writing z as (x, y), consider the expression

$$E = f(x + h, y + k) - f(x + h, y) - f(x, y + k) + f(x, y) \ ,$$

where (h, k) are so small that $(x + h, y + k)$ belongs to G. Letting

$$E = \varphi(x + h) - \varphi(x) \ , \quad \text{where } \varphi(x) = f(x, y + k) - f(x, y) \ ,$$

φ is differentiable near x. Applying the Mean-Value Theorem to φ, there is some x^* between x and $x + h$ such that

$$E = \varphi'(x^*)h = f_x(x^*, y + k)h - f_x(x^*, y)h \ .$$

As for fixed x^*, f_x is differentiable in the second component near y, we can apply the Mean-Value Theorem once more to get

$$E = f_{xy}(x^*, y^*)hk \ , \qquad \text{some } y^* \text{ between } y \text{ and } y + k \ .$$

On the other hand, write

$$E = \psi(y + k) - \psi(y) \ , \quad \text{where } \psi(y) = f(x + h, y) - f(x, y) \ .$$

Following the same track we arrive at

$$E = f_{yx}(x', y')hk \ , \qquad \text{for some mean values } x', y' \ .$$

By comparing these two expressions for E, we obtain

$$f_{xy}(x^*, y^*)hk = f_{yx}(x', y')kh \ .$$

Taking (h, k) to be a sequence $(h_n, k_n) \to 0$ with $h_n, k_n \neq 0$. After canceling the term $h_n k_n$ and then using the continuity of the second derivatives at (x, y), we arrive at the desired result.

□

A repeated use of this proposition shows that different orders of taking partial derivatives will lead to the same result as long as the partial derivatives under consideration are continuous. In many applications, the functions involved are smooth, that is, they have partial derivatives of arbitrary order, so their derivatives are independent of the order they are taken.

In the single variable case the existence of the derivative at a point implies continuity of the function at this point. But in higher dimension this is not always true. The following example illustrates this disappointing phenomenon.

Example 5.4. Consider the function

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

This function is defined in the entire $\mathbb{R}^2$. It is clear that its partial derivatives exist at every point other than the origin. To see what happens at the origin, we go back to the definition of the derivative. As $f(x,0) = 0$ for all x and $f(0,y) = 0$ for all x and y, we conclude that the partial derivatives of f at the origin exist and equal to 0. However, letting $(x,0)$ tend to $(0,0)$, we see that f tends to 0. On the other hand, letting (x,x) tend to $(0,0)$, f tends to $1/2$. (In fact, $f(x,x)$ is always equal to $1/2$.) Hence f is not continuous at $(0,0)$.

5.2 Differentiability of Functions

Now we come to the notion of differentiability. It should be distinguished from that of the existence of the partial derivatives for $n \geq 2$. Recall that for a function f defined on (a,b) and $x_0 \in (a,b)$, it is called differentiable at x_0 if there exists some α such that

$$\lim_{x \to x_0} \frac{f(x) - f(x_0) - \alpha(x - x_0)}{x - x_0} = 0 \ .$$

The function $l(x) = f(x_0) + \alpha(x - x_0)$ is a linear function. Differentiability is concerned with the quantitative approximation of a function near at a certain point by a linear function. Nevertheless, by comparing with the definition of the derivative, we see differentiability and the existence of derivative are the same. We have

Theorem. *If a function f is differentiable at x_0, then its derivative exists and equals to α. Conversely, if the derivative of f exists at x_0, then f is differentiable at x_0 with α given by $f'(x_0)$.*

Things become quite different in higher dimensions. Let f be defined in some open set G in $\mathbb{R}^n$ and $z \in G$. The function f is called **differentiable** at z if there exists a linear function

$$L(x) = f(z) + \sum_{j=1}^{n} c_j(x_j - z_j) = f(z) + c \cdot (x - z) \ , \quad c = (c_1, \cdots, c_n) \ .$$

satisfying

$$\lim_{x \to z} \frac{|f(x) - L(x)|}{|x - z|} = 0 \ .$$

Here the denominator in the definition is equal to

$$|x - z| = \sqrt{(x_1 - z_1)^2 + \cdots (x_n - z_n)^2} \ .$$

Using the little $\circ$ notation, we can express this definition as

$$f(x) = f(z) + c \cdot (x - z) + \circ(|x - z|) \ .$$

This notation is very handy when it comes to operating limits. Here $\circ(|x - z|)$ denotes a quantity which satisfies $\circ(|x - z|)/|x - z| \to 0$ as $x \to z$. Differentiability shows how the function is approximated by a linear function near a point. Whenever the function is differentiable at z, the linear function

$$L(x) = f(z) + c \cdot (x - z)$$

is called the **differential** or **linear approximation** of f at z. We will see in a minute that c_j must be equal to the j-th partial derivative of f at z whenever f is differentiable at z, so the differential is uniquely specified.

Now we show that differentiability implies continuity in all dimensions.

Theorem 5.2. *If f is differentiable at z, then it is continuous at z.*

Proof. If f is differentiable at z, we have

$$f(x) - f(z) = c \cdot (x - z) + \circ(|x - z|) \ .$$

By Cauchy-Schwarz Inequality,

$$\begin{aligned} \lim_{x \to z} |f(x) - f(z)| &= \lim_{x \to z} |c \cdot (x - z) + \circ(|x - z|)| \\ &\le \lim_{x \to z} |c||x - z| + \lim_{x \to z} \circ(|x - z|) \\ &= 0 \ . \end{aligned}$$

□

Next we study the relation between differentiability and partial differentiation. We first show that differentiability implies the existence of partial derivatives. At this point, we introduce the **gradient vector**

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right) .$$

It is regarded as a map from G to $\mathbb{R}^n$ (provided the partial derivatives exist).

Theorem 5.3. *Let f be defined in some open set $G \in \mathbb{R}^n$ and $z \in G$. Then f is differentiable at z if and only if its partial derivatives exist at z and*

$$f(x) = f(z) + \nabla f(z) \cdot (x - z) + \circ(|x - z|) .$$

Therefore, the differential of f at z is given by $L(x) = f(z) + \nabla f(z) \cdot (x - z)$.

Proof. When f is differentiable at z, let its differential be $L(x) = f(z) + c \cdot (x - z)$. Taking $x = (x_1, z_2, \cdots, z_n)$, $x - z = (x_1 - z_1, 0, \cdots, 0)$ and $|x - z| = |x_1 - z_1|$. We have

$$f(x) - f(z) = c \cdot (x - z) + \circ(|x_1 - z_1|) = c_1(x_1 - z_1) + \frac{\circ(|x_1 - z_1|)}{x_1 - z_1}$$

and

$$\frac{f(x_1, z_2, \cdots, z_n) - f(z_1, z_2, \cdots, z_n)}{x_1 - z_1} = c_1 + \circ(|x_1 - z_1|) .$$

Hence $\partial f / \partial x_1(z)$ exists and equals to c_1. Similarly one shows that $c_j = \partial f / \partial x_j$ for other j's.

Conversely, it is clear from the relation

$$f(x) = f(z) + \nabla f(z) \cdot (x - z) + \circ(|x - z|) ,$$

that f is differentiable at z with the differential as claimed. □

It follows from this theorem that the existence of partial derivatives is a necessary condition for differentiability. But it is not sufficient. In fact, the function in Example 5.4 has partial derivatives at $(0, 0)$ but it is not even continuous there, let alone differentiable!

To establish differentiability via the existence of partial derivatives, one needs to impose more on the function. The following frequently used result ensures that differentiability follows from the continuity of the partial derivatives.

Theorem 5.4. *Suppose that the partial derivatives of f exist in the open set G. Then f is differentiable at $z \in G$ provided all partial derivatives are continuous at z.*

Proof. Let us examine the case $n = 2$. Suppose that the partial derivatives of f exist near (x, y) and is continuous at (x, y). For small (h, k), by the mean-value theorem

$$\begin{aligned} f(x+h, y+k) - f(x, y) &= (f(x+h, y+k) - f(x, y+k)) + (f(x, y+k) - f(x, y)) \\ &= \frac{\partial f}{\partial x}(x^*, y+k)h + \frac{\partial f}{\partial y}(x, y^*)k \ , \end{aligned}$$

where x^* and y^* lie between $x, x+h$ and $y, y+k$ respectively. Write it as

$$f(x+h, y+k) - f(x, y) = \frac{\partial f}{\partial x}(x, y)h + \frac{\partial f}{\partial y}(x, y)k + T \ ,$$

where

$$T = \left(\frac{\partial f}{\partial x}(x^*, y+k) - \frac{\partial f}{\partial x}(x, y) \right) h + \left(\frac{\partial f}{\partial y}(x, y^*) - \frac{\partial f}{\partial y}(x^*, y) \right) k \ .$$

But now the continuity of the partial derivatives at (x, y) shows that $T/|(h, k)| \to 0$ as $(h, k) \to (0, 0)$, so $T = \circ(|(h, k)|)$. We conclude that f is differentiable at (x, y).

In general, write $f(x) - f(z)$ as

$$\begin{aligned} &(f(x_1, x_2, x_3, \cdots, x_n) - f(z_1, x_2, x_3, \cdots, x_n)) + (f(z_1, x_2, x_3, \cdots, x_n) - f(z_1, z_2, x_3, \cdots, x_n)) \\ &+ (f(z_1, z_2, x_3, \cdots, x_n) - f(z_1, z_2, z_3, \cdots, x_n)) + (f(z_1, z_2, z_3, \cdots, x_n) - f(z_1, z_2, z_3, \cdots, z_n)) \ . \end{aligned}$$

and then proceed in a similar way. □

In the following example we display a function which is differentiable at a certain point but whose partial derivatives are not continuous at this point. It shows that continuity of the partial derivatives is a sufficient but not a necessary condition for differentiability.

Example 5.5. Study the differentiability of the function

$$h(x, y) = xy \sin \frac{1}{x^2 + y^2} \ , \quad (x, y) \neq (0, 0) \ ,$$

and $h(0, 0) = 0$. We claim that it is differentiable at $(0, 0)$. We will prove this claim by verifying the definition directly. In fact, using $h_x(0, 0) = h_y(0, 0) = 0$, the differential of h at $(0, 0)$ must be 0, and

$$\begin{aligned} \left| \frac{h(x, y) - 0}{\sqrt{x^2 + y^2}} \right| &= \left| \frac{xy \sin 1/(x^2 + y^2)}{\sqrt{x^2 + y^2}} \right| \\ &\le \frac{|xy|}{\sqrt{x^2 + y^2}} \\ &\le \frac{\sqrt{x^2 + y^2}}{2} \to 0 \ , \quad \text{as } |(x, y)| = \sqrt{x^2 + y^2} \to 0 \ , \end{aligned}$$

which shows that h is differentiable at $(0,0)$. On the other hand,

$$h_x(x,y) = y\sin\frac{1}{x^2+y^2} - \frac{2x^2y}{(x^2+y^2)^2}\cos\frac{1}{x^2+y^2} \ .$$

In particular,

$$h_x(x,x) = x\sin\frac{1}{2x^2} - \frac{1}{2x}\cos\frac{1}{2x^2} \ ,$$

which is clearly not continuous at $(0,0)$.

Now we study the properties of differentiability. Just like continuity, differentiability is preserved under basic algebraic operations. We have learned this for functions of a single variable, The situation is the same in the multi-variable case.

Theorem 5.5. *Let f and g be defined in some open G and $z \in G$. Suppose that f and g are differentiable at z. Then for $\alpha, \beta \in \mathbb{R}$,*

(a) $\alpha f + \beta g$ is differentiable at z,

(b) fg is differentiable at z,

(c) f/g is differentiable at z provided $g(z) \neq 0$.

Proof. By assumption we have

$$f(x) = f(z) + c\cdot(x-z) + T_1, \text{ and } g(x) = g(z) + d\cdot(x-z) + T_2 \ ,$$

where $c = \nabla f(z), d = \nabla g(z)$ and $T_i = \circ(|x-z|),\ i = 1,2,$ as $x \to z$. It follows that

$$\alpha f(x) + \beta g(x) = \alpha f(z) + \beta g(z) + (\alpha c + \beta d)\cdot(x-z) + T_3 \ ,$$

where

$$T_3 = \alpha T_1 + \beta T_2 \ .$$

As it is obvious that $T_3 = \circ(|x-z|)$ as $x \to z$, (a) holds.

Next, we have

$$f(x)g(x) = f(z)g(z) + (g(z)c + f(z)d)\cdot(x-z) + T_4 \ ,$$

where

$$T_4 = (f(z) + c\cdot(x-z))\,T_2 + (g(z) + d\cdot(x-z))\,T_1 + c\cdot(x-z)d\cdot(x-z) + T_1T_2 \ .$$

Again $T_4 = \circ(|x-z|)$ as $x \to z$, so (b) holds.

Finally, as f/g can be written as fg^{-1}, in view of (b), (c) follows if we can show that $1/g$ is differentiable when $g(z) \neq 0$. Indeed,

$$\begin{aligned}\frac{1}{g(x)} - \frac{1}{g(z)} &= \frac{-1}{g(x)g(z)}\left(g(x) - g(z)\right) \\ &= \frac{-1}{g(x)g(z)}\left(d \cdot (x-z) + T_2\right) \\ &= \frac{-d}{g(z)^2} \cdot (x-z) + T_5 \ ,\end{aligned}$$

where

$$T_5 = \left(\frac{-d}{g(z)g(x)} + \frac{d}{g(z)^2}\right) \cdot (x-z) - \frac{T_2}{g(x)g(z)} \ .$$

As $T_5 = \circ(|x-z|)$ as $x \to z$, $1/g$ is differentiable at z.

□

Theorem 5.6. *A polynomial in smooth everywhere. A rational function is smooth in its natural domain.*

Proof. As straightforward from definition, the linear function

$$x \mapsto \sum_{j=1}^{n} a_j x_j + b \ , \quad a_j, b \in \mathbb{R}, \quad x = (x_1, \cdots, x_n) \ ,$$

is differentiable everywhere. Using the addition and product rules in Theorem 5.5, we see that all polynomials are smooth. On the other hand, for a rational function $p(x)/q(x)$, its natural domain is given by the open set $\{x \in \mathbb{R}^n : \ q(x) \neq 0\}$. Theorem 5.5(c) ensures that it is smooth in this set.

□

Theorem 5.4 and Theorem 5.5 together guarantee differentiability for many functions far beyond the rational ones. Before giving you any examples, let us review the differentiability property of the elementary functions. In last chapter we recalled their natural domains and established their continuity over these domains. Since differentiability is a more stringent property, the domain of differentiability could be a proper subset of the natural domain. Now we summarize their differentiability as follows.

- The power function $t \mapsto t^a, a \geq 1$, is continuously differentiable on $[0, \infty)$ and smooth on $(0, \infty)$.
- The radical function $t \mapsto t^a, a \in (0, 1)$ is continuous on $[0, \infty)$ and smooth on $(0, \infty)$.
- The exponential function $t \mapsto e^t$ is smooth on $(-\infty, \infty)$.
- The logarithmic function $t \mapsto \log t$ is smooth on $(0, \infty)$.

- The trigonometric functions $\sin t, \cos t$ is smooth on $(-\infty, \infty)$ and $\tan t$ is smooth on $(-\infty, \infty) \setminus \{(n + \frac{1}{2})\pi, n \in \mathbb{Z}\}$.
- The inverse trigonometric functions $\arcsin t, \arccos t$, and $\arctan t$ are smooth in their respective domains.
- The absolute value function $t \mapsto |t|$ is continuous on $(-\infty, \infty)$ and smooth on $(-\infty, 0) \cup (0, \infty)$. It is not differentiable at 0.

A function is **continuously differentiable** if its partial derivatives are continuous. It is **smooth** if its derivatives of all order exist.

Example 5.6. Discuss the differentiability of the following functions:

(a) $f(x, y) = \sqrt{x^2 + y^2}$,

(b) $g(x, y) = \log(1 - x^2 - y^2)$, and

(c) $h(x, y) = e^{xy} \log(1 - x^2 - y^2)$.

(a) The square function is continuous on $[0, \infty)$ and smooth on $(0, \infty)$. Therefore f is a continuous function in $\mathbb{R}^2$. For $(x, y) \neq (0, 0)$, $x \mapsto \sqrt{x^2 + y^2}$ is differentiable by the chain rule in calculus and $f_x(x, y) = x/\sqrt{x^2 + y^2}$ is continuous in $\mathbb{R}^2$ except at $(0, 0)$. Similarly, $y \mapsto \sqrt{x^2 + y^2}$ is differentiable and $f_y(x, y) = y/\sqrt{x^2 + y^2}$ is continuous. By Theorem 5.4, f is differentiable in $\mathbb{R}^2 \setminus \{(0, 0)\}$. On the other hand, note that $f(x, 0) = |x|$ which is not differentiable at 0, so $f_x(0, 0)$ does not exist. This fact suffices in showing that f is not differentiable at $(0, 0)$.

(b) The logarithmic function is smooth in $(0, \infty)$. Therefore, the natural domain of g is the open unit disk $D = \{(x, y) : x^2 + y^2 < 1\}$. We have

$$\frac{\partial g}{\partial x} = \frac{-2x}{1 - x^2 - y^2}, \quad \frac{\partial g}{\partial y} = \frac{-2y}{1 - x^2 - y^2} \quad .$$

These are rational functions restricted to D. By Theorem 5.4 g is differentiable in D.

(c) The function h is the product of e^{xy} and g. From $\dfrac{\partial e^{xy}}{\partial x} = ye^{xy}$ and $\dfrac{\partial e^{xy}}{\partial y} = xe^{xy}$, we see both partial derivatives are continuous. Hence e^{xy} is differentiable everywhere. As the product of two differentiable functions, Theorem 5.5 tells us that h is differentiable everywhere.

When Theorems 5.4 and 5.5 fail to apply, differentiability has to be examined by going back to definition as in Example 5.5.

In last chapter it was shown that compositions of continuous functions are still continuous. It is natural to wonder whether compositions of differentiable functions are still differentiable. A positive answer will be given by the chain rule in the next section.

5.3 The Chain Rules

We start with a special case.

Theorem 5.7 (**Chain Rule I**). *Let*

(a) $f : G \subset \mathbb{R}^n \to \mathbb{R}$*, where* G *is open, be differentiable at* $z \in G$*, and*

(b) $\Phi : I \to \mathbb{R}$*, where* I *is an open interval, be differentiable at* $f(z) \in I$*.*

Then $\Phi \circ f$ *is differentiable at* z *and*

$$\frac{\partial \Phi \circ f}{\partial x_j}(z) = \Phi'(f(z))\frac{\partial f}{\partial x_j}(z) \ , \quad j = 1, \cdots n.$$

Proof. By assumption there is a small ball B around z such that

$$f(x) = f(z) + \nabla f(z) \cdot (x - z) + \circ(|x - z|), \quad \forall x \in B \ ,$$

as $x \to z$. On the other hand, since Φ is differentiable at t_0,

$$\Phi(t) - \Phi(t_0) = \Phi'(t_0)(t - t_0) + \circ(|t - t_0|).$$

Taking $t = f(x)$ and $t_0 = f(z)$, we have

$$\begin{aligned} \Phi(f(x)) - \Phi(f(z)) &= \Phi'(f(z))(f(x) - f(z)) + \circ(|f(x) - f(z)|) \\ &= \Phi'(f(z)) \left(\nabla f(z) \cdot (x - z) + \circ(|x - z|)\right) + \circ(|f(x) - f(z)|) \\ &= \Phi'(f(z))\nabla f(z) \cdot (x - z) + \circ(|x - z|) \ , \end{aligned}$$

after observing that

$$\Phi'(f(z)) \circ (|x - z|) = \circ(|x - z|) \ ,$$

and

$$\circ(|f(x) - f(z)|) = \circ \left(|\nabla f(z) \cdot (x - z)| + \circ(|x - z|)\right) = \circ(|x - z|) \ .$$

□

Example 5.7. Discuss the differentiability of the following functions:

(a) $f(x,y) = \sqrt{x^2+y^2}$,

(b) $g(x,y) = \log(1-x^2-y^2)$, and

This is contained in Theorem 5.4. We re-examine it in light of Chain Rule I. Indeed, f is the composition of the polynomial x^2+y^2 which is smooth from $\mathbb{R}^2$ to $[0,\infty)$ and $\sqrt{z}$ is smooth from $(0,\infty)$ to $\mathbb{R}$. It follows immediately from Chain Rule I that f is differentiable in $\mathbb{R}^2 \setminus \{(0,0)\}$. Similarly, g is differentiable in $\{(x,y) : x^2+y^2 < 1\}$ as it is the composition of two differentiable functions $(x,y) \mapsto 1-x^2-y^2$ on $\{(x,y) : x^2+y^2 < 1\}$ and $z \to \log z,\ z \in (0,\infty)$.

When it comes to the differentiability of a composite function $g \circ f$ where g is smooth and f is differentiable but with discontinuous partial derivatives, Chain Rule I applies to establish differentiability. However, Theorem 5.4 cannot be used. It shows that the chain rule is not only more theoretically satisfying but also applies to a wider range.

Now let us formulate a general chain rule. In Chapter 7 we will encounter an even more general chain rule for vector-valued functions. Although the setting is getting more and more involved, the proof is basically the same.

Let $F : G \mapsto \mathbb{R}^m$ where $G \subset \mathbb{R}^n$ is open. Writing $F(x) = (f_1(x), \cdots, f_m(x))$, F is called continuous (resp. differentiable) if each $f_j, j = 1, \cdots, m$, is **continuous** (resp. **differentiable**).
jn

Theorem 5.8 (Chain Rule II). *Let*

(a) $F : G \subset \mathbb{R}^n \to \mathbb{R}^m$*, where* G *is open, be differentiable at* $z \in G$*, and*

(b) $\Phi : G_1 \subset \mathbb{R}^m \to \mathbb{R}$*, where* G_1 *is open, be differentiable at* $F(z) \in G_1$*.*

Then $\Phi \circ F : G \to \mathbb{R}$ *is differentiable at* z *and*

$$\frac{\partial \Phi \circ F}{\partial x_j}(z) = \sum_{k=1}^{m} \frac{\partial \Phi}{\partial u_k}(F(z)) \frac{\partial f_k}{\partial x_j}(z) \ , \quad k = 1, \cdots, n.$$

Consequently, when both F *and* Φ *are continuously differentiable in their respective domains,* $\Phi \circ F$ *is continuously differentiable.*

Proof. This rule can be obtained from Chain Rule I by looking at each component of the function. Indeed, since Φ is differentiable at some u_0,

$$\Phi(u) - \Phi(u_0) = \sum_{j=1}^{m} \frac{\partial \Phi}{\partial u_j}(u_0)(u_j - u_{0j}) + \circ(|u-u_0|) \ .$$

On the other hand, as F is differentiable at z,

$$f_j(x) - f_j(z) = \sum_{k=1}^{n} \frac{\partial f_j}{\partial x_k}(z)(x_k - z_k) + \circ(|x - z|) \ , \quad j = 1, \cdots, m.$$

Letting $u = F(x)$ and $u_0 = F(z)$ and combining these two expressions yields

$$\begin{aligned} \Phi(F(x)) - \Phi(F(z)) &= \sum_{j=1}^{m} \frac{\partial \Phi}{\partial u_j}(F(z))(f_j(x) - f_j(z)) + \circ(|F(x) - F(z)|) \\ &= \sum_{j,k} \frac{\partial \Phi}{\partial u_j}(F(z)) \frac{\partial f_j}{\partial x_k}(z)(x_k - z_k) + T \ , \end{aligned}$$

where

$$T = \sum_{j=1}^{m} \frac{\partial \Phi}{\partial u_j}(F(z)) \circ (|x - z|) + \circ(|F(x) - F(z)|) \ .$$

Using

$$\left| \sum_{j=1}^{m} \frac{\partial \Phi}{\partial u_j}(F(z)) \circ (|x - z|) \right| = \circ(|x - z|) \ ,$$

and

$$|F(x) - F(z)| \leq \left| \sum_{k=1}^{n} \frac{\partial F}{\partial x_k}(z)(x_k - z_k) + \circ(|x - z|) \right| \leq C|x - z| \ ,$$

for some constant C, hence $T = \circ(|x - z|)$ and

$$\Phi(F(x)) - \Phi(F(z)) = \sum_{j,k} \frac{\partial \Phi}{\partial u_j}(F(z)) \frac{\partial f_j}{\partial x_k}(z)(x_k - z_k) + \circ(|x - z|) \ .$$

We conclude that $\Phi \circ F$ is differentiable at z with the given first order approximation as claimed.

$\square$

Example 5.8. Find the partial derivatives of the function $g \circ \gamma$ where $\gamma(t) = (t^3, \sin t^2)$ and $g(u, v) = uv + e^u$. We have $\gamma : \mathbb{R} \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}$, so the composite function $g \circ \gamma : \mathbb{R} \to \mathbb{R}$. By Chain Rule II,

$$\begin{aligned} \frac{dg \circ \gamma}{dt}(t) &= \frac{\partial g}{\partial u}(\gamma(t))\gamma_1'(t) + \frac{\partial g}{\partial v}(\gamma(t))\gamma_2'(t) \\ &= (v + e^u)(\gamma(t)) \times 3t^2 + u(\gamma(t)) \times 2t \cos t^2 \\ &= 3(\sin t^2 + e^{t^3})t^2 + 2t^4 \cos t^2 \ . \end{aligned}$$

Example 5.9. Find M_x, M_y, M_{xy} for $M(xy, x^2 - y^2, \log x)$. More precisely, here we have $F(x,y) = (f_1, f_2, f_3)(x,y) = (xy, x^2 - y^2, \log x)$. We are asked to find the partial derivatives of the composite function $N(x,y) = M \circ F(x,y)$. Here the setting is $F : \mathbb{R}^2 \to \mathbb{R}^3$ and $M : \mathbb{R}^3 \to \mathbb{R}$ so $N : \mathbb{R}^2 \to \mathbb{R}$. We have

$$\begin{aligned} \frac{\partial N}{\partial x} &= \frac{\partial M}{\partial u}\frac{\partial f_1}{\partial x} + \frac{\partial M}{\partial v}\frac{\partial f_2}{\partial x} + \frac{\partial M}{\partial w}\frac{\partial f_3}{\partial x} \\ &= \frac{\partial M}{\partial u} y + \frac{\partial M}{\partial v}(2x) + \frac{\partial M}{\partial w}\frac{1}{x} \\ &= y\frac{\partial M}{\partial u} + 2x\frac{\partial M}{\partial v} + \frac{1}{x}\frac{\partial M}{\partial w} \ . \end{aligned}$$

$$\begin{aligned} \frac{\partial N}{\partial y} &= \frac{\partial M}{\partial u}\frac{\partial f_1}{\partial y} + \frac{\partial M}{\partial v}\frac{\partial f_2}{\partial y} + \frac{\partial M}{\partial w}\frac{\partial f_3}{\partial y} \\ &= x\frac{\partial M}{\partial u} - 2y\frac{\partial M}{\partial v} \ . \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 N}{\partial x \partial y} &= \frac{\partial M}{\partial u} + x\left(y\frac{\partial^2 M}{\partial u^2} + 2x\frac{\partial^2 M}{\partial v \partial u} + \frac{1}{x}\frac{\partial^2 M}{\partial w \partial u}\right) \\ &\quad -2y\left(\frac{\partial^2 M}{\partial u \partial v} + 2x\frac{\partial^2 M}{\partial v^2} + \frac{1}{x}\frac{\partial^2 M}{\partial w \partial v}\right) \\ &= \frac{\partial M}{\partial u} + xy\frac{\partial^2 M}{\partial u^2} + 2(x^2 - y)\frac{\partial^2 M}{\partial u \partial v} + \frac{\partial^2 M}{\partial u \partial w} - 4xy\frac{\partial^2 M}{\partial v^2} - \frac{2y}{x}\frac{\partial^2 M}{\partial v \partial w} \ . \end{aligned}$$

We have assumed the partial derivatives are independent of the order of differentiation. In many books, the same notation M will be used for $N = M \circ F$. It is understood in the context, for instance, in M_u it means regarding M as a function of u, v while in M_x it means regarding M as a function of $N = M \circ F$, that is, N_x .

The chain rule is widely applied in study of differential equations. We will illustrate how it is applied in two ways.

First, it is used in transforming partial differentiable equations. Let us show how to express the two dimensional Laplace equation

$$\Delta f = 0 \ ,$$

where Δ is the Laplace operator given by

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \ ,$$

in the polar coordinates. For $f = f(x,y)$, let

$$\tilde{f}(r,\theta) = f(x,y) \ ,$$

where the variables are related by

$$x = r\cos\theta, \quad y = r\sin\theta \ ,$$

or

$$r = \sqrt{x^2+y^2}, \quad \tan\theta = \frac{y}{x} \ .$$

We would like to see what equation $\tilde{f}$ satisfies when f solves the Laplace equation.

By the Chain Rule,

$$\frac{\partial f}{\partial x} = \frac{\partial \tilde{f}}{\partial r}\frac{\partial r}{\partial x} + \frac{\partial \tilde{f}}{\partial \theta}\frac{\partial \theta}{\partial x} \ ,$$

$$\frac{\partial^2 f}{\partial x^2} = \left(\frac{\partial^2 \tilde{f}}{\partial r^2}\frac{\partial r}{\partial x} + \frac{\partial^2 \tilde{f}}{\partial\theta\partial r}\frac{\partial \theta}{\partial x}\right)\frac{\partial r}{\partial x} + \frac{\partial \tilde{f}}{\partial r}\frac{\partial^2 r}{\partial x^2} + \left(\frac{\partial^2 \tilde{f}}{\partial r\partial\theta}\frac{\partial r}{\partial x} + \frac{\partial^2 \tilde{f}}{\partial\theta^2}\frac{\partial \theta}{\partial x}\right)\frac{\partial \theta}{\partial x} + \frac{\partial \tilde{f}}{\partial \theta}\frac{\partial^2 \theta}{\partial x^2} \ .$$

$$\frac{\partial f}{\partial y} = \frac{\partial \tilde{f}}{\partial r}\frac{\partial r}{\partial y} + \frac{\partial \tilde{f}}{\partial \theta}\frac{\partial \theta}{\partial y} \ ,$$

$$\frac{\partial^2 f}{\partial y^2} = \left(\frac{\partial^2 \tilde{f}}{\partial r^2}\frac{\partial r}{\partial y} + \frac{\partial^2 \tilde{f}}{\partial\theta\partial r}\frac{\partial \theta}{\partial y}\right)\frac{\partial r}{\partial y} + \frac{\partial \tilde{f}}{\partial r}\frac{\partial^2 r}{\partial y^2} + \left(\frac{\partial^2 \tilde{f}}{\partial r\partial\theta}\frac{\partial r}{\partial y} + \frac{\partial^2 \tilde{f}}{\partial\theta^2}\frac{\partial \theta}{\partial y}\right)\frac{\partial \theta}{\partial y} + \frac{\partial \tilde{f}}{\partial \theta}\frac{\partial^2 \theta}{\partial y^2} \ .$$

We have

$$\frac{\partial r}{\partial x} = \frac{x}{r} \ , \quad \frac{\partial r}{\partial y} = \frac{y}{r} \ ,$$

$$\frac{\partial \theta}{\partial x} = \frac{-\sin\theta}{r} \ , \quad \frac{\partial \theta}{\partial y} = \frac{\cos\theta}{r} \ ,$$

$$\frac{\partial^2 r}{\partial x^2} = \frac{y^2}{r^3} \ , \quad \frac{\partial^2 r}{\partial y^2} = \frac{x^2}{r^3} \ ,$$

$$\frac{\partial^2 \theta}{\partial x^2} = \frac{2\sin\theta\cos\theta}{r^2} \ , \quad \frac{\partial^2 \theta}{\partial y^2} = \frac{-2\sin\theta\cos\theta}{r^2} \ .$$

Plug in to get

$$\Delta f(x,y) = \left(\frac{\partial^2 \tilde{f}}{\partial r^2} + \frac{1}{r}\frac{\partial \tilde{f}}{\partial r} + \frac{1}{r^2}\frac{\partial^2 \tilde{f}}{\partial \theta^2}\right)(r,\theta) \ .$$

We conclude that the Laplace operator is

$$\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} \ ,$$

in the polar coordinates. The Laplace equation in the polar coordinates for $g = g(r, \theta)$ is now given by

$$\frac{\partial^2 g}{\partial r^2} + \frac{1}{r}\frac{\partial g}{\partial r} + \frac{1}{r^2}\frac{\partial^2 g}{\partial \theta^2} = 0 \ .$$

Solutions to the Laplace equation are called **harmonic functions**. Harmonic functions describe equilibrium or steady states. They comes up in mathematics, physics, engineering, and many other places.

You may wonder why we express the equation in the polar coordinates. An immediate reward is some special solutions can be found easily. For instance, assuming the solution $g = g(r)$ is independent of θ, the equation reduces to

$$\frac{\partial^2 g}{\partial r^2} + \frac{1}{r}\frac{\partial g}{\partial r} = 0 \ ,$$

which is readily solved to give $g(r) = \log r$. We conclude that the function $\log(x^2 + y^2)$ is a harmonic function in the rectangular coordinates. Similarly, a function independent of r satisfies $g_{\theta\theta} = 0$, hence $g(\theta) = \theta$ is a solution. Writing back in the rectangular coordinates, it shows that $\arctan y/x$ is a harmonic function.

Again in practise people do not distinguish between f and $\tilde{f}$. It will be clear from the context. For instance, when f_r is present it means $\tilde{f}_r$.

The second application of the chain rule is to find special solutions of PDE's. In the situation above, in principle, any solution in (x, y) can be expressed as a solution in (r, θ) and vice versa. Here we are not so ambitious; we are settled to find special solutions, that is, solutions assuming form certain forms. In many cases, these special solutions are either fundamental ones or provides insight for the study.

Example 5.10. Consider the Korteweg-de Vries (KdV) equation

$$u_t + u_{xxx} + uu_x = 0 \ ,$$

which describes waves in shallow water. First, we look for traveling waves, that is, special solutions of the form $u(x, t) = \varphi(x - ct)$. Plugging in the equation, we see that φ must satisfy the simpler equation

$$\varphi_{yyy} + \varphi\varphi_y - c\varphi_y = 0 \ ,$$

or

$$\varphi_{yy} + \frac{1}{2}\varphi^2 - c\varphi = k \ ,$$

after one integration. A second integration reduces its order further by one:

$$\frac{1}{2}\varphi_y^2 = -\frac{1}{6}\varphi^3 + \frac{c}{2}\varphi^2 + k\varphi + l, \quad k, l \in \mathbb{R} \ .$$

One can verify that the function

$$\varphi(y) = 3c \operatorname{sech}^2\left(\frac{\sqrt{c}}{2}y\right),$$

solves this equation with $k = l = 0$. The special solution $u(x,t) = \varphi(x - ct)$ of the KdV equation is called a soliton.

Next, we look for solution of the form $u(x,t) = t^{-2/3}\psi(t^{-1/3}x)$. From

$$u_x = t^{-1}\psi_y, \quad u_{xxx} = t^{-5/3}\psi_{yyy}, \quad u_t = \frac{-1}{3}t^{-5/3}(y\psi_y + 2\psi) ,$$

we see that ψ must satisfy

$$\psi_{yyy} + \psi\psi_y - \frac{1}{3}y\psi_y - \frac{2}{3}\psi = 0 .$$

The solution for the KdV equation $u(x,t)$ depends on two variables and now the special solution $\psi(y)$ depends on one variable only. Although the equation it satisfies is still difficult to solve, it is simpler than the original one.

Not arbitrary form can be reduced. For instance, letting $u(x,t) = \eta(x/t)$, the equation becomes

$$\frac{1}{t^2}\eta_{yyy} - y\eta_{yy} + \eta\eta_y = 0 .$$

There is no way to get rid of the variable t resulting in a single differential equation for η. The special forms from which the equation can be reduced rely on the symmetries of the differential equation, a topic out of scope.

5.4 Directional Derivatives and The Gradient

Let f be defined in an open G and $x \in G$. Let ξ be any direction, that is, $\xi \in \mathbb{R}^n, |\xi| = 1$. We define the **directional derivative** of f at x along the direction ξ to be the derivative

$$D_\xi f(x) = \lim_{t \to 0} \frac{f(x + t\xi) - f(x)}{t} ,$$

provided the limit exists. Note that in particular,

$$D_{e_j} f = \frac{\partial f}{\partial x_j} , \quad j = 1, \cdots, n.$$

Theorem 5.9. *Setting as above, the directional derivative of f exists along every direction at a differentiable point and*

$$D_\xi f(x) = \xi \cdot \nabla f(x), \quad \forall \xi \neq (0, \cdots, 0) .$$

Proof. As f is differentiable at x, for all sufficiently small t,

$$\begin{aligned} f(x+t\xi)-f(x) &= \nabla f(x)\cdot(x+t\xi-x)+\circ(|x+t\xi-x|) \\ &= t\nabla f(x)\cdot\xi+\circ(|t|)\ ,\ \text{as } t\to 0\ . \end{aligned}$$

Dividing this relation by t and then letting t tend to 0, the theorem follows. □

Recall that Cauchy-Schwarz Inequality asserts that for $a, b \in \mathbb{R}^n$, $|a \cdot b| \leq |a||b|$ with equality holds if and only if a and b are proportional to each other. Whenever the gradient vector is non-zero, $\nabla f(x)/|\nabla f(x)|$ becomes a direction. Applying Cauchy-Schwarz Inequality,

$$-|\nabla f(x)| \leq |\xi \cdot \nabla f(x)| \leq |\nabla f(x)|,$$

and equality holds if and only if ξ is $\pm\nabla f(x)/|\nabla f(x)|$. In other words, among all directions, $D_\xi f(x)$ becomes the maximum at $\xi = \nabla f(x)/|\nabla f(x)|$ and the minimum at $\xi = -\nabla f(x)/|\nabla f(x)|$. This interesting interpretation tells us that the gradient direction is the direction along which the function increases most rapidly, and the negative direction is the direction along which the function decreases most rapidly. Imagine the graph of a function is the landscape of a mountain. To go to the top as quickly as possible one should follow the direction of the gradient at each point.

Example 5.11. Consider the function

$$H(x,y,z) = x^2 - y^2 + 2xz\ , \quad (x,y,z) \in \mathbb{R}^3\ .$$

Find

(a) its directional derivative along $\xi = (4,3,0)/5$ at $P(-1,2,3)$,

(b) the direction it increases most rapidly at P, and

(c) the direction it decreases most rapidly at P.

The gradient of H is

$$\nabla H(x,y,z) = (2x+2z, -2y, 2x)\ ,$$

so $\nabla H = (4,-4,-2)$ at P.

(a) By definition,

$$D_\xi H = \xi \cdot \nabla H = \frac{1}{5}(4,3,0)\cdot(4,-4,-2) = \frac{4}{5}\ .$$

(b) The direction of most rapid increase is the direction of the gradient

$$\frac{(4,-4,-2)}{|(4,-4,-2)|} = \frac{1}{3}(2,-2,-1)\ .$$

(c) The direction of most rapid decrease is $\frac{1}{3}(-2,2,1)$.

5.5 Differentials as Approximation

From the definition of differentiability we see that whenever a function is differentiable at a point,

$$f(x) - (f(z) + \nabla f(z) \cdot (x - z)) = \circ(|x - z|) \ .$$

As $\circ(|x-z|)$ becomes even smaller when $|x-z|$ is small, we could ignor it and use the differential

$$L(x) = f(z) + \nabla f(z) \cdot (x - z)$$

to approximate $f(x)$ for x near z.

The error, that is, the difference between the actual value and the approximate value, can be estimated when more information on $\circ(|x-z|)$ is known. The following theorem gives some ideas on the effective range of the approximation. It can be skipped in a first reading.

Theorem 5.10. ** Assume that f admits continuous partial derivatives up to second order in $B_r(z)$. Then*

$$|f(x) - (f(z) + \nabla f(z) \cdot (x - z))| \leq \frac{M}{2}|x - z|^2 \ ,$$

where M is a bound on the second derivatives over $B_r(z)$.

The differential

$$L(dx) = f(x) + \nabla f(x) \cdot dx,$$

where x is fixed, is now a function of the variables $dx = (dx_1, \cdots, dx_n)$, where the notation dx is used to emphasis that it is a small quantity. We also let

$$df = \nabla f(x) \cdot dx$$

to represent the approximate error. Comparing with the exact error Δf, we have

$$\Delta f = f(x + dx) - f(x) \ ,$$

and

$$df = L(df) - f(x) \ .$$

So the difference between the exact and approximate errors is given by $\Delta f - df$ which is of order $\circ(|dx|)$.

Example 5.12. Find an approximate value of $\sqrt{1.02^3 + 1.93^3}$. We choose

$$f(x, y) = \sqrt{x^3 + y^3},$$

and take $(x, y) = (1, 2)$ and $(dx, dy) = (0.02, -0.07)$. The differential at $(1, 2)$ is given by

$$\begin{aligned} & f(1,2) + f_x(1,2)(1.02-1) + f_y(1,2)(1.93-2) \\ = & f(1,2) + f_x(1,2) \times 0.02 + f_y(1,2) \times (-0.07) \ . \end{aligned}$$

Using

$$f_x = \frac{3}{2}\frac{x^2}{\sqrt{x^3+y^3}}, \quad \text{and } f_y = \frac{3}{2}\frac{y^2}{\sqrt{x^3+y^3}},$$

$f_x(1, 2) = 1/2$ and $f_y(1, 2) = 2$. Therefore, the approximate value of $f(1.02, 1.93)$ is given by the differential

$$\sqrt{9} + \frac{1}{2} \times 0.02 - 2 \times 0.07 = 2.88 \ .$$

One may use Theorem 5.10 to estimate the difference between this approximate value with the accurate value. We will not do it here.

Example 5.13. The volume of a cylinder is given by the formula

$$V = \pi r^2 h \ ,$$

where r is the radius of the base disk and h its height. Now r is measured with an error up to 1% and h up to 0.5%. What is the induced error on the volume? Here $V = V(r, h)$ is a function of r and h. For small changes in r and h, we may use the approximate change

$$dV = 2\pi rhdr + \pi r^2 dh$$

as a good approximation of the real change ΔV, which is given by

$$\pi(r+dr)^2(h+dh) - \pi r^2 h \ .$$

Using $dr = \pm 0.01r$ and $dh = 0.005$,

$$dV = \pm 2\pi rh \times \pm 0.01r + \pi r^2 \times \pm 0.005h = \pm 0.0225 \times V \ .$$

We conclude that the error in volume would be up to 0.0255, that is, 2.55%.

The formula for the volume is quite simple, so we can carry out the error estimate in an accurate way. Indeed,

$$\begin{aligned} \Delta V - dV &= 2\pi r dr dh + \pi(dr)^2 h + \pi(dr)^2 dh \\ &= 2\pi r \times (\pm 0.01r) \times (\pm 0.005h) + \pi(\pm 0.01r)^2 h + \pi(\pm 0.01r)^2 \times (\pm 0.005h) \ . \end{aligned}$$

We conclude

$$\frac{|\Delta V - dV|}{V} \le 0.0002005 = 0.02\% \ .$$

Comments on Chapter 5.

This chapter is the core of this course. You should understand it well. In particular, pay attention to the following topics:

- The relation between partial derivatives and differentiability (Theorems 5.3 and 5.4).
- The differentiability of a function either by Theorems 5.5 and 5.6 or going back to the definition.
- Master the use of the chain rule especially in transforming differential equations.
- The meaning of the gradient of a function.
- Calculating approximate errors of functions.

Supplementary Readings

3.1, 3.2, 3.3, 3.5, 4.1–4.4 in [Au]. 14.3, 14.5 (excluding tangent planes), and 14.6 in [Thomas]. We have separated the geometric aspect of functions from its analytic aspect. In particular, tangent planes will be covered in next chapter.