Calculus 微積分

Ching-hsiao Arthur Cheng 鄭經斅

Chapter 13

Functions of Several Variables

13.1 Introduction to Functions of Several Variables

Definition 13.1

Let D be a set of ordered pairs of real numbers. If to each ordered pair (x, y) in D there corresponds a unique real number $f(x, y)$, then f is a real-valued function of (two variables) x and y. The set D is the domain of f, and the corresponding set of values for $f(x, y)$ is the range of f. For the function $z = f(x, y)$, x and y are called the independent variables and z is called the dependent variable.

Definition 13.2

Let f, g be real-valued functions of two variables with domain D.

1. The sum of f and g, the difference of f and g and the product of f and g, denoted by $f + g$, $f - g$ and fg, are functions defined on D given by

$$\begin{aligned} (f+g)(x,y) &= f(x,y) + g(x,y) \quad && \forall\,(x,y) \in D\,, \\ (f-g)(x,y) &= f(x,y) - g(x,y) \quad && \forall\,(x,y) \in D\,, \\ (fg)(x,y) &= f(x,y)g(x,y) \quad && \forall\,(x,y) \in D\,. \end{aligned}$$

2. The quotient of f and g, denoted by $\dfrac{f}{g}$, is a function defined on $D \backslash \big\{(x, y) \in D \,\big|\, g(x, y) = 0\big\}$ given by

$$\frac{f}{g}(x,y) = \frac{f(x,y)}{g(x,y)} \qquad \forall\,(x,y) \in D \text{ such that } g(x,y) \neq 0\,.$$

Remark 13.3. A function f of two variables should be given along with its domain. When the domain of a function is not specified, as before the domain should be treated as the collection of all (x, y) such that $f(x, y)$ is meaningful.

Definition 13.4

Let h be a real-valued function of two variables with domain D, and $g : I \to \mathbb{R}$ be a real-valued function (of one variable) on an interval I. The composite function of g and h, denoted by $g \circ h$, is a function defined on $D \cap \{(x, y) \in D \,|\, h(x, y) \in I\}$ given by

$$(g \circ h)(x, y) = g\big(h(x, y)\big) \qquad \forall\, (x, y) \in D \text{ such that } h(x, y) \in I\,.$$

Similar concepts such as real-valued functions of three variables, the sum, different, product, quotient and composition of functions of three variables can be defined accordingly.

Definition 13.5

Let D be a set of ordered pairs of real numbers, and $f : D \to \mathbb{R}$ be a real-valued function of two variables. The graph of f is the set of all points (x, y, z) for which $z = f(x, y)$ and $(x, y) \in D$.

Example 13.6. Let $r > 0$ be a real number. The graph of the function $z = f(x, y) = \sqrt{r^2 - x^2 - y^2}$ is the upper hemi-sphere of the sphere centered at the origin with radius r. On the other hand, the graph of the function $z = g(x, y) = -\sqrt{r^2 - x^2 - y^2}$ is the lower hemi-sphere of the sphere.

Definition 13.7: Level Curves

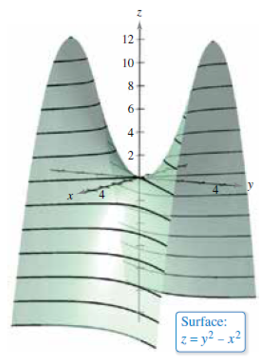
Let D be a set of ordered pairs of real numbers, and $f : D \to \mathbb{R}$ be a function of two variables. A level curve (or contour curve) of f is a collection of points (x, y) in D along which the value of $f(x, y)$ is constant.

Definition 13.8: Level Surfaces

Let D be a set of ordered pairs of real numbers, and $f : D \to \mathbb{R}$ be a function of three variables. A level surface of f is a collection of points (x, y, z) in D along which the value of $f(x, y, z)$ is constant.

Example 13.9. A level curve of the function $z = \sqrt{r^2 - x^2 - y^2}$ is a circle centered at the origin, and a level surface of the function $w = g(x, y, z) = x^2 + y^2 + z^2 - r^2$ is a sphere centered at the origin.

Example 13.10. The graph of $f(x, y) = y^2 - x^2$ is called a hyperbolic paraboloid. A level curve of a hyperbolic paraboloid is a hyperbola (or degenerated hyperbola), and each plane perpendicular to the xy-plane intersects the graph of $z = f(x, y)$ along a parabola (or degenerated parabola).



Hyperbolic paraboloid

13.2 Limits and Continuity

Definition 13.11

Let $\delta > 0$ be given. The δ-neighborhood about a point (x_0, y_0) in the plane is the open disk centered at (x_0, y_0) with radius δ given by

$$D\big((x_0, y_0), \delta\big) \equiv \Big\{(x, y) \,\Big|\, \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta\Big\} .$$

Definition 13.12

Let R be a collection of points in the plane. A point (x_0, y_0) (in R) is called an ***interior point*** of R if there exists $\delta > 0$ such that the δ-neighborhood about (x_0, y_0) lies entirely in R. If every point in R is an interior point of R, then R is called an open region. A point (x_0, y_0) is called a ***boundary point*** of R if every δ-neighborhood about (x_0, y_0) containing points inside R and point outsides R. In other words, (x_0, y_0) is a boundary point of R if

$$\forall\, \delta > 0, D\big((x_0, y_0), \delta\big) \cap R \neq \varnothing \quad \text{and} \quad D\big((x_0, y_0), \delta\big) \cap R^{\complement} \neq \varnothing\,.$$

If R contains all its boundary points, then R is called a closed region.

Remark 13.13. For $x \in \mathbb{R}$ and $\delta > 0$, let $D(x, \delta)$ denote the interval $(x - \delta, x + \delta)$ (and called the interval centered at x with radius r). Then for each $x \in (a, b)$, there exists $\delta > 0$ such that $D(x, r) \subseteq (a, b)$; thus (a, b) is called an open interval. The end-points a, b of the interval are boundary points of the interval, and $[a, b]$ is a closed interval since it contains all its boundary points.

Definition 13.14

Let f be a real-valued function of two variables defined, except possibly at (x_0, y_0), on an open disk centered at (x_0, y_0), and let L be a real number. Then

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

if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\big|f(x) - L\big| < \varepsilon \quad \text{whenever} \quad 0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta\,.$$

Remark 13.15. If $\lim\limits_{(x,y)\to(x_0,y_0)} f(x, y) = L_1$ and $\lim\limits_{(x,y)\to(x_0,y_0)} f(x, y) = L_2$, then $L_1 = L_2$. In other words, the limit is unique when it exists.

The proof of the following is almost identical to the one of Theorem 1.14.

Theorem 13.16: Properties of Limits of Functions of Two Variables

Let $(a, b) \in \mathbb{R}^2$. Suppose that the limits

$$\lim_{(x,y)\to(a,b)} f(x, y) = L \qquad \text{and} \qquad \lim_{(x,y)\to(a,b)} g(x, y) = K\,.$$

both exist, and c is a constant.

1. $\lim\limits_{(x,y)\to(a,b)} c = c$, $\lim\limits_{(x,y)\to(a,b)} x = a$ and $\lim\limits_{(x,y)\to(a,b)} y = b$.

2. $\lim\limits_{(x,y)\to(a,b)} \big[f(x, y) \pm g(x, y)\big] = L + K$;

3. $\lim\limits_{(x,y)\to(a,b)} \big[f(x, y)g(x, y)\big] = LK$;

4. $\lim\limits_{(x,y)\to(a,b)} \dfrac{f(x, y}{g(x, y)} = \dfrac{L}{K}$ if $K \neq 0$.

Theorem 13.17: Squeeze

Let $(x_0, y_0) \in \mathbb{R}^2$. Suppose that f, g, h are functions of two variables such that

$$g(x, y) \leqslant f(x, y) \leqslant h(x, y)$$

except possible at (x_0, y_0), and $\lim\limits_{(x,y)\to(x_0,y_0)} g(x, y) = \lim\limits_{(x,y)\to(x_0,y_0)} h(x, y) = L$, then

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

Example 13.18. For $(a, b) \in \mathbb{R}^2$, find the limit $\lim\limits_{(x,y)\to(a,b)} \dfrac{5x^2y}{x^2 + y^2}$.

First we note that 1-3 of Theorem 13.16 implies that the function $f(x, y) = 5x^2y$ and $g(x, y) = x^2 + y^2$ has the property that

$$\lim_{(x,y)\to(a,b)} f(x, y) = 5a^2b \quad \text{and} \quad \lim_{(x,y)\to(a,b)} g(x, y) = a^2 + b^2\,.$$

Therefore, Theorem 13.16 again shows the following:

1. If $(a, b) \neq (0, 0)$, then 4 of Theorem 13.16 implies that

$$\lim_{(x,y)\to(a,b)} \frac{5x^2y}{x^2 + y^2} = \lim_{(x,y)\to(a,b)} \frac{f(x, y)}{g(x, y)} = \frac{5a^2b}{a^2 + b^2}\,.$$

2. If $(a, b) = (0, 0)$, then we cannot apply 4 of Theorem 13.16 to compute the limit. Nevertheless, since

$$\Big|\frac{5x^2y}{x^2 + y^2} - 0\Big| \leqslant 5|y| \qquad \forall\,(x, y) \neq (0, 0)\,,$$

the Squeeze Theorem implies that

$$\lim_{(x,y)\to(0,0)} \frac{5x^2y}{x^2 + y^2} = 0\,.$$

Example 13.19. Show that the limit $\lim\limits_{(x,y)\to(0,0)} \Big(\dfrac{x^2 - y^2}{x^2 + y^2}\Big)^2$ does not exist.

Let $f(x, y) = \Big(\dfrac{x^2 - y^2}{x^2 + y^2}\Big)^2$. By the definition of limits, if $\lim\limits_{(x,y)\to(0,0)} f(x, y) = L$ exists, then there exists $\delta > 0$ such that

$$\big|f(x, y) - L\big| < \frac{1}{2} \quad \text{whenever} \quad 0 < \sqrt{x^2 + y^2} < \delta$$

which implies that

$$L - \frac{1}{2} < f(x,y) < L + \frac{1}{2} \quad \text{whenever} \quad 0 < \sqrt{x^2+y^2} < \delta\,. \tag{13.2.1}$$

However, when (x,y) satisfies $0 < \sqrt{x^2+y^2} < \delta$ and $x = y$, then $f(x,y) = 0$ while on the other hand, when (x,y) satisfies $0 < \sqrt{x^2+y^2} < \delta$ and $y = 0$, then $f(x,y) = 1$. This is a contradiction because of (13.2.1).

• Another way of looking at this limit: When (x,y) approaches $(0,0)$ along the line $x = y$ (we use the notation $\lim\limits_{\substack{(x,y)\to(0,0)\\ x=y}}$ to denote this limit process), we find that

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

and when (x,y) approaches $(0,0)$ along the x-axis (we use the notation $\lim\limits_{\substack{(x,y)\to(0,0)\\ y=0}}$ to denote this limit process), we find that

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

The uniqueness of the limit implies that the limit of f at $(0,0)$ does not exist.

13.2.1 Continuity of functions of two variables

Definition 13.20

A function f of two variables is ***continuous at a point*** (x_0,y_0) in an open region R if $f(x_0,y_0)$ is defined and is equal to the limit of $f(x,y)$ as (x,y) approaches (x_0,y_0); that is,

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

In other words, f is continuous at (x_0,y_0) if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\big|f(x,y) - f(x_0,y_0)\big| < \varepsilon \quad \text{whenever} \quad \sqrt{(x-x_0)^2+(y-y_0)^2} < \delta\,.$$

The function f is ***continuous in the open region*** R if it is continuous at every point in R.

Remark 13.21. 1. Unlike the case that f does not have to be defined at (x_0,y_0) in order to consider the limit of f at (x_0,y_0), for f to be continuous at a point (x_0,y_0) f has to be defined at (x_0,y_0).

2. A point (x_0, y_0) is called a discontinuity of f if f is not continuous at (x_0, y_0). (x_0, y_0) is called a ***removable discontinuity*** of f if $\lim\limits_{(x,y)\to(x_0,y_0)} f(x, y)$ exists.

Theorem 13.22

Let f and g be functions of two variables such that f and g are continuous at (x_0, y_0).

1. $f \pm g$ is continuous at (x_0, y_0).
2. fg is continuous at (x_0, y_0).
3. $\dfrac{f}{g}$ is continuous at (x_0, y_0) if $g(x_0, y_0) \neq 0$.

Theorem 13.23

If h is continuous at (x_0, y_0) and g is continuous at $h(x_0, y_0)$, then the composite function $g \circ h$ is continuous at (x_0, y_0); that is,

$$\lim_{(x,y)\to(x_0,y_0)} (g \circ h)(x, y) = g\big(h(x_0, y_0)\big) .$$

13.3 Partial Derivatives

Definition 13.24

Let f be a function of two variable. The first partial derivative of f with respect to x at (x_0, y_0), denoted by $f_x(x_0, y_0)$, is defined by

$$f_x(x_0, y_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}$$

provided the limit exists. The first partial derivative of f with respect to y at (x_0, y_0), denoted by $f_y(x_0, y_0)$, is defined by

$$f_y(x_0, y_0) = \lim_{\Delta y \to 0} \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}$$

provided the limit exists. When f_x and f_y exist for all (x_0, y_0) (in a certain open region), f_x and f_y are simply called the first partial derivative of f with respect to x and y, respectively.

• **Notation**: For $z = f(x, y)$, the partial derivative f_x and f_y, can also be denoted by

$$\frac{\partial}{\partial x} f(x, y) = f_x(x, y) = z_x = \frac{\partial z}{\partial x} = \frac{\partial f}{\partial x}(x, y)$$

and

$$\frac{\partial}{\partial y} f(x, y) = f_y(x, y) = z_y = \frac{\partial z}{\partial y} = \frac{\partial f}{\partial y}(x, y)\,.$$

When evaluating the partial derivative at (x_0, y_0), we write

$$f_x(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial}{\partial x}\Big|_{(x,y)=(x_0,y_0)} f(x, y)$$

and

$$f_y(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = \frac{\partial}{\partial y}\Big|_{(x,y)=(x_0,y_0)} f(x, y)\,.$$

Example 13.25. For $f(x, y) = xe^{x^2 y}$, find f_x and f_y, and evaluate each at the point $(1, \ln 2)$.

Note that f_x is obtained by treating y as a constant and differentiate f with respect to x. Therefore, the product rule implies tat

$$f_x(x, y) = \Big(\frac{\partial}{\partial x} x\Big) e^{x^2 y} + x\Big(\frac{\partial}{\partial x} e^{x^2 y}\Big) = e^{x^2 y} + x \cdot e^{x^2 y} \cdot 2xy = (1 + 2x^2 y)e^{x^2 y}\,;$$

thus

$$f_x(1, \ln 2) = (1 + 2\ln 2)e^{\ln 2} = 2(1 + 2\ln 2)\,.$$

Similarly,

$$f_y(x, y) = \Big(\frac{\partial}{\partial y} x\Big) e^{x^2 y} + x\Big(\frac{\partial}{\partial y} e^{x^2 y}\Big) = x^3 e^{x^2 y}\,;$$

thus $f_y(1, \ln 2) = e^{\ln 2} = 2$.

Example 13.26. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

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

Then if $(x, y) \neq (0, 0)$, we can apply the quotient rule (and product rule) to compute the partial derivatives and obtain that

$$\begin{aligned} f_x(x, y) &= \frac{(x^2 + y^2)\frac{\partial}{\partial x}\big[xy(x^2 - y^2)\big] - xy(x^2 - y^2)\frac{\partial}{\partial x}(x^2 + y^2)}{(x^2 + y^2)^2} \\ &= \frac{(x^2 + y^2)\big[y(x^2 - y^2) + 2x^2 y\big] - xy(x^2 - y^2)\cdot(2x)}{(x^2 + y^2)^2} \\ &= \frac{x^4 y + 4x^2 y^3 - y^5}{(x^2 + y^2)^2}\,. \end{aligned}$$

If $(x, y) = (0, 0)$, we cannot use the quotient rule to compute the derivative since the denominate is 0 (so that 4 of Theorem 13.16 cannot be applied), and we have to compute $f_x(0, 0)$ using the definition. By definition,

$$f_x(0,0) = \lim_{\Delta x \to 0} \frac{f(\Delta x, 0) - f(0,0)}{\Delta x} = 0\,.$$

Therefore,

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

Similarly,

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

• **Geometric meaning of partial derivatives**: Let $f(x, y)$ be a function of two variable, (x_0, y_0) be given, and $z_0 = f(x_0, y_0)$. Consider the graph of the function $z = f(x, y_0)$ (of one variable) on the xz-plane. If the graph $z = f(x, y_0)$ has a tangent line at (x_0, z_0), then the slope of the tangent line at (x_0, z_0) is given by

$$\lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

and this limit, if exists, is $f_x(x_0, y_0)$. This is called ***the slopes in the x-direction of the surface** $z = f(x, y)$ **at the point*** (x_0, y_0, z_0). Similarly, the slope of the tangent line of the graph of $z = f(x_0, y)$ at (y_0, z_0) is $f_y(x_0, y_0)$, and is called ***the slopes in the y-direction of the surface** $z = f(x, y)$ **at the point*** (x_0, y_0, z_0).

• **Partial derivatives of functions of three or more variables:**

The concept of partial derivatives can be extended to functions of three or more variables. For example, if $w = f(x, y, z)$, then

$$\begin{aligned} \frac{\partial w}{\partial x} &= f_x(x,y,z) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y, z) - f(x,y,z)}{\Delta x}\,, \\ \frac{\partial w}{\partial y} &= f_y(x,y,z) = \lim_{\Delta y \to 0} \frac{f(x, y+\Delta y, z) - f(x,y,z)}{\Delta y}\,, \\ \frac{\partial w}{\partial z} &= f_z(x,y,z) = \lim_{\Delta z \to 0} \frac{f(x, y, z+\Delta z) - f(x,y,z)}{\Delta z}\,. \end{aligned}$$

In general, if $w = f(x_1, x_2, \cdots, x_n)$, then there are n first partial derivatives denoted by

$$\frac{\partial w}{\partial x_k} = f_{x_k}(x_1, x_2, \cdots, x_n)\,, \quad k = 1, 2, \cdots, n\,.$$

• **Higher-order partial derivatives:**

We can also take higher-order partial derivatives of functions of several variables. For example, for $z = f(x, y)$,

1. Differentiate twice with respect to x:
$$\frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial x}\Big) = \frac{\partial^2 f}{\partial x^2} = f_{xx}\,.$$
2. Differentiate twice with respect to y:
$$\frac{\partial}{\partial y}\Big(\frac{\partial f}{\partial y}\Big) = \frac{\partial^2 f}{\partial y^2} = f_{yy}\,.$$
3. Differentiate first with respect to x and then with respect to y:
$$\frac{\partial}{\partial y}\Big(\frac{\partial f}{\partial x}\Big) = \frac{\partial^2 f}{\partial y \partial x} = f_{xy}\,.$$
4. Differentiate first with respect to y and then with respect to x:
$$\frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial y}\Big) = \frac{\partial^2 f}{\partial x \partial y} = f_{yx}\,.$$

The third and fourth cases are called ***mixed partial derivatives***.

Example 13.27. In this example we compute the second partial derivatives of the function given in 13.26. We have obtained that

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

and

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

If $(x, y) \neq (0, 0)$, the quotient rule, the product rule and the chain rule (for functions of one variable) together show that

$$\begin{aligned} f_{xx}(x, y) &= \frac{(x^2+y^2)^2 \frac{\partial}{\partial x}(x^4y + 4x^2y^3 - y^5) - (x^4y + 4x^2y^3 - y^5)\frac{\partial}{\partial x}(x^2+y^2)^2}{(x^2+y^2)^4} \\ &= \frac{(x^2+y^2)^2(4x^3y + 8xy^3) - (x^4y + 4x^2y^3 - y^5)\cdot\big[2(x^2+y^2)\cdot(2x)\big]}{(x^2+y^2)^3} \\ &= \frac{(x^2+y^2)(4x^3y + 8xy^3) - 4x(x^4y + 4x^2y^3 - y^5)}{(x^2+y^2)^3} = \frac{-4x^3y^3 + 12xy^5}{(x^2+y^2)^3}\,. \end{aligned}$$

Similarly, if $(x, y) \neq (0,0)$,

$$f_{yy}(x,y) = \frac{(x^2+y^2)^2(-8x^3y - 4xy^3) - (x^5 - 4x^3y^2 - xy^4) \cdot \left[2(x^2+y^2) \cdot (2y)\right]}{(x^2+y^2)^2}$$

$$= \frac{-12x^5y + 4x^3y^3}{(x^2+y^2)^3},$$

$$f_{xy}(x,y) = \frac{(x^2+y^2)(x^4+12x^2y^2-5y^4) - 4y(x^4y+4x^2y^3-y^5)}{(x^2+y^2)^3}$$

$$= \frac{x^6+9x^4y^2-9x^2y^4-y^6}{(x^2+y^2)^3}$$

and

$$f_{yx}(x,y) = \frac{(x^2+y^2)(5x^4-12x^2y^2-y^4) - 4x(x^5-4x^3y^2-xy^4)}{(x^2+y^2)^3}$$

$$= \frac{x^6+9x^4y^2-9x^2y^4-y^6}{(x^2+y^2)^3}.$$

We note that when $(x,y) \neq (0,0)$, $f_{xy}(x,y) = f_{yx}(x,y)$.

Since $f_x(x,0) = f_y(0,y) = 0$ for all $x \neq 0$, we find that

$$f_{xx}(0,0) = \lim_{\Delta x \to 0} \frac{f_x(\Delta x, 0) - f_x(0,0)}{\Delta x} = 0$$

and

$$f_{yy}(0,0) = \lim_{\Delta y \to 0} \frac{f_y(0, \Delta y) - f_y(0,0)}{\Delta y} = 0.$$

Finally, we compute $f_{xy}(0,0)$ and $f_{yx}(0,0)$. By definition,

$$f_{xy}(0,0) = \lim_{\Delta y \to 0} \frac{f_x(0,\Delta y) - f_x(0,0)}{\Delta y} = \lim_{\Delta y \to 0} \frac{\frac{-\Delta y^5}{\Delta y^4}}{\Delta y} = -1$$

and

$$f_{yx}(0,0) = \lim_{\Delta x \to 0} \frac{f_y(\Delta x, 0) - f_y(0,0)}{\Delta x} = \lim_{\Delta y \to 0} \frac{\frac{\Delta x^5}{\Delta x^4}}{\Delta x} = 1.$$

We note that $f_{xy}(0,0) \neq f_{yx}(0,0)$.

Theorem 13.28: Clairaut's Theorem

If f is a function of x and y such that f_{xy} and f_{yx} are continuous on an open disk D, then

$$f_{xy}(x,y) = f_{yx}(x,y) \qquad \forall\,(x,y) \in D.$$

In the following, we prove the following more general version:

If f is a function of x and y such that on an open disk D f_{xy} is continuous and f_{yx} exists, then $f_{xy}(x,y) = f_{yx}(x,y)$ for all $(x,y) \in D$.

Proof. Let $(a,b) \in D$ be given. Then

$$
\begin{aligned}
f_{yx}(a,b) = (f_y)_x(a,b) &= \lim_{h\to 0} \frac{f_y(a+h,b) - f_y(a,b)}{h} \\
&= \lim_{h\to 0} \frac{\displaystyle\lim_{k\to 0} \frac{f(a+h,b+k) - f(a+h,b)}{k} - \lim_{k\to 0} \frac{f(a,b+k) - f(a,b)}{k}}{h} \\
&= \lim_{h\to 0}\lim_{k\to 0} \frac{f(a+h,b+k) - f(a,b+k) - f(a+h,b) - f(a,b)}{hk} .
\end{aligned}
$$

Define

$$
Q(h,k) \equiv \frac{f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)}{hk} .
$$

Then the computation above shows that

$$
\lim_{h\to 0}\lim_{k\to 0} Q(h,k) = f_{yx}(a,b) . \tag{13.3.1}
$$

For $h,k \neq 0$ such that $(a+h,b+k) \in D$, define $\varphi(x,y) = f(x,y+k) - f(x,y)$. Then $Q(h,k) = \dfrac{\varphi(a+h,b) - \varphi(a,b)}{hk}$. By the mean value theorem for functions of one variable (Theorem 3.9),

$$
Q(h,k) = \frac{\varphi_x(a+\theta_1 h,b)h}{hk} = \frac{f_x(a+\theta_1 h,b+k) - f_x(a+\theta_1 h,b)}{k}
$$

for some functions $\theta_1 = \theta_1(h)$ satisfying $0 < \theta_1 < 1$. Applying the mean value theorem again,

$$
\begin{aligned}
Q(h,k) &= \frac{f_x(a+\theta_1 h,b+k) - f_x(a+\theta_1 h,b)}{k} = \frac{f_{xy}(a+\theta_1 h,b+\theta_2 k)k}{k} \\
&= f_{xy}(a+\theta_1 h,b+\theta_2 k)
\end{aligned}
$$

for some functions $\theta_2 = \theta_2(h,k)$ satisfying $0 < \theta_2 < 1$. Therefore, we establish that there exist functions $\theta_1 = \theta_1(h)$ and $\theta_2 = \theta_2(h,k)$ such that

$$
Q(h,k) = f_{xy}(x+\theta_1 h, y+\theta_2 k) .
$$

Passing to the limit as $k \to 0$ first then $h \to 0$, using (13.3.1) and the continuity of f_{xy} we conclude that $f_{xy}(a,b) = f_{yx}(a,b)$. □

Example 13.29. Let $f(x,y,z) = ye^x + x\ln z$. Then $f_x(x,y,z) = ye^x + \ln z$, $f_y(x,y,z) = e^x$ and $f_z(x,y,z) = \dfrac{x}{z}$. Therefore,

$$\begin{aligned}
f_{xy}(x,y,z) &= e^x = f_{yx}(x,y,z)\,,\\
f_{xz}(x,y,z) &= \frac{1}{z} = f_{zx}(x,y,z) \quad \forall\, z \neq 0\,,\\
f_{yz}(x,y,z) &= 0 = f_{zy}(x,y,z)\,.
\end{aligned}$$

13.4 Differentiability of Functions of Several Variables

Recall that a function $f : (a,b) \to \mathbb{R}$ is said to be differentiable at a point $c \in (a,b)$ if the limit

$$\lim_{\Delta x \to 0} \frac{f(c+\Delta x) - f(c)}{\Delta x}$$

exists. The differentiability of f at c can be rephrased as follows:

A function $f : (a,b) \to \mathbb{R}$ is said to be differentiable at $c \in (a,b)$ if there exists $m \in \mathbb{R}$ such that

$$\lim_{\Delta x \to 0} \left| \frac{f(c+\Delta x) - f(c) - m\Delta x}{\Delta x} \right| = 0\,.$$

or equivalently,

$$\lim_{x \to c} \left| \frac{f(x) - f(c) - m(x-c)}{x - c} \right| = 0\,.$$

This equivalent way of defining differentiability of functions of one variable motivate the following

Definition 13.30

Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $f : R \to \mathbb{R}$ be a function of two variables. For $(x_0, y_0) \in R$, f is said to be differentiable at (x_0, y_0) if there exist real numbers A, B such that

$$\lim_{(x,y)\to(x_0,y_0)} \frac{\big|f(x,y) - f(x_0,y_0) - (A,B)\cdot(x-x_0, y-y_0)\big|}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} = 0\,.$$

Suppose that f is differentiable at (x_0, y_0). When (x,y) approaches (x_0, y_0) along the line $y = y_0$, we find that

$$0 = \lim_{\substack{(x,y)\to(x_0,y_0)\\ y=y_0}} \frac{\big|f(x,y) - f(x_0,y_0) - A(x-x_0) - B(y-y_0)\big|}{\sqrt{(x-x_0)^2+(y-y_0)^2}}$$
$$= \lim_{x\to x_0} \frac{\big|f(x,y_0) - f(x_0,y_0) - A(x-x_0)\big|}{|x-x_0|} = \lim_{x\to x_0} \Big|\frac{f(x,y_0)-f(x_0,y_0)}{x-x_0} - A\Big|$$

which implies that the number A must be $f_x(x_0,y_0)$. Similarly, $B = f_y(x_0,y_0)$, and we have the following alternative

Definition 13.31

Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $f : R \to \mathbb{R}$ be a function of two variables. For $(x_0,y_0) \in R$, f is said to be differentiable at (x_0,y_0) if ($f_x(x_0,y_0)$, $f_y(x_0,y_0)$ both exist and)

$$\lim_{(x,y)\to(x_0,y_0)} \frac{\big|f(x,y) - f(x_0,y_0) - f_x(x_0,y_0)(x-x_0) - f_y(x_0,y_0)(y-y_0)\big|}{\sqrt{(x-x_0)^2+(y-y_0)^2}} = 0\,.$$

Remark 13.32. The ordered pair $\big(f_x(x_0,y_0), f_y(x_0,y_0)\big)$ if called the derivative of f at (x_0,y_0) if f is differentiable at (x_0,y_0) and is usually denoted by $(Df)(x_0,y_0)$.

2. Using ε-δ notation, we find that f is differentiable at (x_0,y_0) if for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\begin{aligned}&\big|f(x,y) - f(x_0,y_0) - f_x(x_0,y_0)(x-x_0) - f_y(x_0,y_0)(y-y_0)\big| \\ &\qquad \leqslant \varepsilon\sqrt{(x-x_0)^2+(y-y_0)^2} \qquad \text{whenever} \quad \sqrt{(x-x_0)^2+(y-y_0)^2} < \delta\,.\end{aligned}$$

Now suppose that f is a function of two variables such that $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ exist. Define

$$\varepsilon(x,y) = \begin{cases} \dfrac{f(x,y) - f(x_0,y_0) - f_x(x_0,y_0)(x-x_0) - f_y(x_0,y_0)(y-y_0)}{\sqrt{(x-x_0)^2+(y-y_0)^2}} & \text{if } (x,y) \neq (x_0,y_0)\,, \\ 0 & \text{if } (x,y) = (x_0,y_0)\,. \end{cases}$$

Let $\Delta x = x - x_0$, $\Delta y = y - y_0$ and $\Delta z = f(x,y) - f(x_0,y_0)$. Then

$$\Delta z = f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + \varepsilon(x,y)\sqrt{\Delta x^2 + \Delta y^2}\,,$$

and f is differentiable at (x_0,y_0) if and only if $\displaystyle\lim_{(x,y)\to(x_0,y_0)} \varepsilon(x,y) = 0$.

Finally, define

$$
\varepsilon_1(x,y) = \begin{cases} \dfrac{\varepsilon(x,y)\Delta x}{\sqrt{\Delta x^2 + \Delta y^2}} & \text{if } (x,y) \neq (x_0,y_0)\,, \\ 0 & \text{if } (x,y) \neq (x_0,y_0)\,, \end{cases}
\qquad
\varepsilon_2(x,y) = \begin{cases} \dfrac{\varepsilon(x,y)\Delta y}{\sqrt{\Delta x^2 + \Delta y^2}} & \text{if } (x,y) \neq (x_0,y_0)\,, \\ 0 & \text{if } (x,y) \neq (x_0,y_0)\,, \end{cases},
$$

then

$$
0 \leqslant |\varepsilon_1(x,y)|, |\varepsilon_2(x,y)| \leqslant |\varepsilon(x,y)| = \sqrt{\varepsilon_1(x,y)^2 + \varepsilon_2(x,y)^2}
$$

thus the Squeeze Theorem shows that

$$
\lim_{(x,y)\to(x_0,y_0)} \varepsilon(x,y) = 0 \quad \text{if and only if} \quad \lim_{(x,y)\to(x_0,y_0)} \varepsilon_1(x,y) = \lim_{(x,y)\to(x_0,y_0)} \varepsilon_2(x,y) = 0\,.
$$

By the fact that $\varepsilon(x,y)\sqrt{\Delta x^2 + \Delta y^2} = \varepsilon_1(x,y)\Delta x + \varepsilon_2(x,y)\Delta y$, the alternative definition above can be rewritten as

> Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $f : R \to \mathbb{R}$ be a function of two variables. For $(x_0,y_0) \in R$, f is said to be differentiable at (x_0,y_0) if ($f_x(x_0,y_0)$, $f_y(x_0,y_0)$ both exist and) there exist functions ε_1 and ε_2 such that
>
> $$\Delta z = f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y\,,$$
>
> where both ε_1 and ε_2 approaches 0 as $(x,y) \to (x_0,y_0)$.

Example 13.33. Show that the function $f(x,y) = x^2 + 3y$ is differentiable at every point in the plane.

Let $(a,b) \in \mathbb{R}^2$ be given. Then $f_x(a,b) = 2a$ and $f_y(a,b) = 3$. Therefore,

$$
\begin{aligned}
\Delta z - f_x(a,b)\Delta x - f_y(a,b)\Delta y &= x^2 + 3y - a^2 - 3b - 2a(x-a) - 3(y-b) \\
&= (x-a)^2 = \varepsilon_1(x,y)\Delta x + \varepsilon_2(x,y)\Delta y\,,
\end{aligned}
$$

where $\varepsilon_1(x,y) = x - a$ and $\varepsilon_2(x,y) = 0$. Since

$$
\lim_{(x,y)\to(a,b)} \varepsilon_1(x,y) = 0 \qquad \text{and} \qquad \lim_{(x,y)\to(a,b)} \varepsilon_2(x,y) = 0\,,
$$

by the definition we find that f is differentiable at (a,b).

Example 13.34. The function f given in Example 13.26 is differentiable at $(0,0)$ since if $(x,y) \neq (0,0)$,

$$\frac{\big|f(x,y) - f(0,0) - f_x(0,0)x - f_y(0,0)y\big|}{\sqrt{x^2+y^2}} = \frac{\big|xy(x^2-y^2)\big|}{(x^2+y^2)^{\frac{3}{2}}} \leqslant \frac{|x^2-y^2|}{\sqrt{x^2+y^2}} \leqslant |x| + |y|$$

and the Squeeze Theorem shows that

$$\lim_{(x,y)\to(0,0)} \frac{\big|f(x,y) - f(0,0) - f_x(0,0)(x-0) - f_y(0,0)(y-0)\big|}{\sqrt{x^2+y^2}} = 0\,.$$

- **Differentiability of functions of several variables**

A real-valued function f of n variables is differentiable at $(a_1, a_2, \cdots, a_n)$ if there exist n real numbers $A_1, A_2, \cdots, A_n$ such that

$$\lim_{(x_1,\cdots,x_n)\to(a_1,\cdots,a_n)} \frac{\big|f(x_1,\cdots,x_n) - f(a_1,\cdots,a_n) - (A_1,\cdots,A_n)\cdot(x_1-a_1,\cdots,x_n-a_n)\big|}{\sqrt{(x_1-a_1)^2+\cdots+(x_n-a_n)^2}} = 0\,.$$

We also note that when f is differentiable at $(a_1, \cdots, a_n)$, then these numbers $A_1, A_2, \cdots, A_n$ must be $f_{x_1}(a_1, \cdots, a_n)$, $f_{x_2}(a_1, \cdots, a_n)$, $\cdots$, $f_{x_n}(a_1, \cdots, a_n)$, respectively.

It is usually easier to compute the partial derivatives of a function of several variables than determine the differentiability of that function. Is there any connection between some specific properties of partial derivatives and the differentiability? We have the following

Theorem 13.35

Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $f : R \to \mathbb{R}$ be a function of two variables. If f_x and f_y are continuous in a neighborhood of $(x_0, y_0) \in R$, then f is differentiable at (x_0, y_0). In particular, if f_x and f_y are continuous on R, then f is differentiable on R; that is, f is said to be differentiable at every point in R.

Therefore, the differentiability of f in Example 13.26 at any point $(x_0, y_0) \neq (0,0)$ can be guaranteed since f_x and f_y are continuous on $\mathbb{R}^2 \backslash \{(0,0)\}$.

Theorem 13.36

Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $f : R \to \mathbb{R}$ be a function of two variables. If f is differentiable at (x_0, y_0), then f is continuous at (x_0, y_0).

Proof. By the definition of differentiability, if f is differentiable at (x_0, y_0), then there exists function ε_1 and ε_2 such that

$$\lim_{(x,y)\to(x_0,y_0)} \varepsilon_1(x,y) = \lim_{(x,y)\to(x_0,y_0)} \varepsilon_w(x,y) = 0$$

and

$$\begin{aligned} f(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) \\ + \varepsilon_1(x,y)(x-x_0) + \varepsilon_2(x,y)(y-y_0)\,. \end{aligned}$$

Then $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = f(x_0,y_0)$. □

Example 13.37. Consider the function

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

Then f is not continuous at $(0,0)$ since

$$\lim_{\substack{(x,y)\to(0,0)\\ y=0}} f(x,y) = 0 \qquad \text{but} \qquad \lim_{\substack{(x,y)\to(0,0)\\ x=y}} f(x,y) = -\frac{3}{2}\,.$$

However, we note that

$$f_x(0,0) = \lim_{\Delta x\to 0} \frac{f(\Delta x,0) - f(0,0)}{\Delta x} = 0 \quad \text{and} \quad f_y(0,0) = \lim_{\Delta y\to 0} \frac{f(0,\Delta y) - f(0,0)}{\Delta y} = 0\,.$$

Therefore, the existence of partial derivatives at a point in all directions does **not** even imply the continuity.

13.5 Chain Rules for Functions of Several Variables

Recall the chain rule for functions of one variable:

> Let I, J be open intervals, $f : J \to \mathbb{R}$, $g : I \to \mathbb{R}$ be real-valued functions, and the range of g is contained in J. If g is differentiable at $c \in I$ and f is differentiable at $g(c)$, then $f \circ g$ is differentiable at c and
>
> $$\frac{d}{dx}\Big|_{x=c} (f\circ g)(x) = f'(g(c))g'(c)\,.$$

For functions of two variables, we have the following

Theorem 13.37

Let $z = f(x, y)$ be a differentiable function (of x and y). If $x = g(t)$ and $y = h(t)$ are differentiable functions (of t), then $z(t) = f\big(x(t), y(t)\big)$ is differentiable and

$$z'(t) = f_x\big(x(t), y(t)\big)x'(t) + f_y\big(x(t), y(t)\big)y'(t)\,.$$

Let $\gamma(t) = \big(x(t), y(t)\big)$. Then $\gamma'(t) = \big(x'(t), y'(t)\big)$, and the chain rule above can be written as

$$\frac{d}{dt}(f \circ \gamma)(t) = (Df)(\gamma(t)) \cdot \gamma'(t)\,.$$

A short-hand notation of the identity above

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = (f_x, f_y) \cdot (x', y')\,.$$

Corollary 13.38

Let $z = f(x, y)$ be a differentiable function (of x and y).

1. If $x = u(s, t)$ and $y = v(s, t)$ are such that $\dfrac{\partial u}{\partial s}$ and $\dfrac{\partial v}{\partial s}$ exist, then the first partial derivative $\dfrac{\partial z}{\partial s}$ of the function $z(s, t) = f\big(u(s, t), v(s, t)\big)$ exists and

$$z_s(s, t) = f_x\big(u(s, t), v(s, t)\big)u_s(s, t) + f_y\big(u(s, t), v(s, t)\big)v_s(s, t)\,.$$

2. If $x = u(s, t)$ and $y = v(s, t)$ are such that $\dfrac{\partial u}{\partial t}$ and $\dfrac{\partial v}{\partial t}$ exist, then the first partial derivative $\dfrac{\partial z}{\partial t}$ of the function $z(s, t) = f\big(u(s, t), v(s, t)\big)$ exists and

$$z_t(s, t) = f_x\big(u(s, t), v(s, t)\big)u_t(s, t) + f_y\big(u(s, t), v(s, t)\big)v_t(s, t)\,.$$

Example 13.39. Let $f(x, y) = x^2y - y^2$. Find $\dfrac{dz}{dt}$, where $z(t) = f(\sin t, e^t)$.

1. Since $z(t) = e^t \sin^2 t - e^{2t}$, by the product rule and the chain rule for functions of one variable, we find that

$$z'(t) = \frac{de^t}{dt}\sin^2 t + e^t\frac{d\sin^2 t}{dt} - 2e^{2t} = e^t \sin^2 t + 2e^t \sin t \cos t - 2e^{2t}\,.$$

2. By the chain rule for functions of two variables,

$$\begin{aligned} z'(t) &= \big(f_x(\sin t, e^t), f_y(\sin t, e^t)\big) \cdot \frac{d}{dt}(\sin t, e^t) \\ &= (2xy, x^2 - 2y)\Big|_{(x,y)=(\sin t, e^t)} \cdot (\cos t, e^t) \\ &= (2e^t \sin t, \sin^2 t - 2e^t) \cdot (\cos t, e^t) \\ &= 2e^t \sin t \cos t + e^t \sin^2 t - 2e^{2t} . \end{aligned}$$

Example 13.40. Let $f(x, y) = 2xy$. Find $\dfrac{\partial z}{\partial s}$ and $\dfrac{\partial z}{\partial t}$, where $z(s, t) = f\big(s^2 + t^2, \frac{s}{t}\big)$.

1. Since $z(s, t) = 2(s^2 + t^2)\dfrac{s}{t} = \dfrac{2s^3}{t} + 2st$, by the product rule we find that

$$\frac{\partial z}{\partial s}(s, t) = \frac{6s^2}{t} + 2t \qquad \text{and} \qquad \frac{\partial z}{\partial t}(s, t) = -\frac{2s^3}{t^2} + 2s .$$

2. By the chain rule for functions of two variables,

$$\begin{aligned} \frac{\partial z}{\partial s}(s, t) &= \big(f_x(s^2 + t^2, s/t), f_y(s^2 + t^2, s/t)\big) \cdot \frac{\partial}{\partial s}\big(s^2 + t^2, \frac{s}{t}\big) \\ &= \big(\frac{2s}{t}, 2(s^2 + t^2)\big) \cdot \big(2s, \frac{1}{t}\big) = \frac{4s^2}{t} + \frac{2s^2 + 2t^2}{t} = \frac{6s^2}{t} + 2t \end{aligned}$$

and

$$\begin{aligned} \frac{\partial z}{\partial t}(s, t) &= \big(f_x(s^2 + t^2, s/t), f_y(s^2 + t^2, s/t)\big) \cdot \frac{\partial}{\partial t}\big(s^2 + t^2, \frac{s}{t}\big) \\ &= \big(\frac{2s}{t}, 2(s^2 + t^2)\big) \cdot \big(2t, -\frac{s}{t^2}\big) = 4s - \frac{2s^3 + 2st^2}{t^2} = -\frac{2s^3}{t^2} + 2s . \end{aligned}$$

• **The chain rule for functions of several variables**

Suppose that $w = f(x_1, x_2, \cdots, x_n)$ be a differentiable function (of $x_1, x_2, \cdots, x_n$). If each x_i is a differentiable function of m variables $t_1, t_2, \cdots, t_m$, then

$$\begin{aligned} \frac{\partial w}{\partial t_1} &= \frac{\partial w}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \frac{\partial w}{\partial x_2}\frac{\partial x_2}{\partial t_1} + \cdots + \frac{\partial w}{\partial x_n}\frac{\partial x_n}{\partial t_1} = \sum_{j=1}^{n} \frac{\partial w}{\partial x_j}\frac{\partial x_j}{\partial t_1} , \\ \frac{\partial w}{\partial t_2} &= \frac{\partial w}{\partial x_1}\frac{\partial x_1}{\partial t_2} + \frac{\partial w}{\partial x_2}\frac{\partial x_2}{\partial t_2} + \cdots + \frac{\partial w}{\partial x_n}\frac{\partial x_n}{\partial t_2} = \sum_{j=1}^{n} \frac{\partial w}{\partial x_j}\frac{\partial x_j}{\partial t_2} , \\ &\ \vdots \\ \frac{\partial w}{\partial t_m} &= \frac{\partial w}{\partial x_1}\frac{\partial x_1}{\partial t_m} + \frac{\partial w}{\partial x_2}\frac{\partial x_2}{\partial t_m} + \cdots + \frac{\partial w}{\partial x_n}\frac{\partial x_n}{\partial t_m} = \sum_{j=1}^{n} \frac{\partial w}{\partial x_j}\frac{\partial x_j}{\partial t_m} . \end{aligned}$$

Using the notation of the matrix multiplication,

$$\begin{bmatrix} \dfrac{\partial w}{\partial t_1} & \dfrac{\partial w}{\partial t_2} & \cdots & \dfrac{\partial w}{\partial t_m} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} \dfrac{\partial x_1}{\partial t_1} & \dfrac{\partial x_1}{\partial t_2} & \cdots & \dfrac{\partial x_1}{\partial t_m} \\ \dfrac{\partial x_2}{\partial t_1} & \dfrac{\partial x_2}{\partial t_2} & \cdots & \dfrac{\partial x_2}{\partial t_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial t_1} & \dfrac{\partial x_n}{\partial t_2} & \cdots & \dfrac{\partial x_n}{\partial t_m} \end{bmatrix}.$$

- **Differentiation of determinant functions**

For an $n \times n$ matrix A, let $\mathrm{Cof}(A)$ denote the cofactor matrix of A; that is, the (i, j)-th entry of $\mathrm{Cof}(A)$ is the determinant of the matrix obtained by deleting the i-th row and j-th column of A or

$$\big[\mathrm{Cof}(A)\big]_{ij} = (-1)^{i+j} \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1(j-1)} & a_{1(j+1)} & \cdots & a_{1\mathrm{n}} \\ \vdots & \ddots & & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & a_{(i-1)2} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)\mathrm{n}} \\ a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)\mathrm{n}} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ a_{\mathrm{n}1} & a_{\mathrm{n}2} & \cdots & a_{\mathrm{n}(j-1)} & a_{\mathrm{n}(j+1)} & \cdots & a_{\mathrm{nn}} \end{vmatrix}.$$

Then the determinant of A, using the reductive algorithm, can be computed by

$$\det(A) = \sum_{k=1}^{n} a_{ik} \big[\mathrm{Cof}(A)\big]_{ik} \qquad \forall\, 1 \leqslant i \leqslant n\,. \tag{13.5.1}$$

On the other hand, the determinant of an $n \times n$ matrix $A = [a_{ij}]_{1 \leqslant i,j \leqslant n}$ can be viewed as a real-valued function of n^2 variable:

$$f(a_{11}, a_{12}, \cdots, a_{1n}, a_{21}, a_{22}, \cdots, a_{2n}, a_{31} \cdots, a_{nn}) = \det([\mathrm{a_{ij}}])\,.$$

Since for each $1 \leqslant i \leqslant n$ the (i, k)-th entry of the cofactor matrix $\mathrm{Cof}(A)_{ik}$ is independent of a_{ij} for all $1 \leqslant j, k \leqslant n$, we have $\dfrac{\partial f}{\partial a_{ij}} = \big[\mathrm{Cof}(A)\big]_{ij}$; thus if $a_{ij} = a_{ij}(t)$ is a function of t for all $1 \leqslant i, j \leqslant n$, with $A = A(t) = \big[a_{ij}(t)\big]_{1 \leqslant i,j \leqslant n}$ in mind the chain rule implies that

$$\frac{d}{dt} f\big(a_{11}(t), a_{12}(t), \cdots, a_{nn}(t)\big) = \sum_{i,j=1}^{n} \big[\mathrm{Cof}(A)\big]_{ij} \frac{da_{ij}(t)}{dt}\,. \tag{13.5.2}$$

Let Adj(A) be the transpose of the cofactor matrix, called the adjoint matrix, of A, then (13.5.2) implies that

$$\frac{d}{dt}\det(A) = \sum_{i,j=1}^{n} \big[\mathrm{Adj}(A)\big]_{ji} \frac{da_{ij}}{dt} = \mathrm{tr}\Big(\mathrm{Adj}(A)\frac{dA}{dt}\Big), \tag{13.5.3}$$

where $\mathrm{tr}(M)$ denotes the trace of a square matrix M and $\dfrac{dA}{dt} = \Big[\dfrac{da_{ij}}{dt}\Big]_{1\leqslant i,j\leqslant n}$. In particular, if A is invertible, then $A^{-1} = \dfrac{1}{\det(A)}\mathrm{Adj}(A)$; thus for invertible matrix $A = \big[a_{ij}(t)\big]$, we have

$$\frac{d}{dt}\det(A) = \mathrm{tr}\Big(\det(A)A^{-1}\frac{dA}{dt}\Big) = \det(A)\mathrm{tr}\Big(A^{-1}\frac{dA}{dt}\Big) \tag{13.5.4}$$

or

$$\frac{d}{dt}\ln\big|\det(A)\big| = \mathrm{tr}\Big(A^{-1}\frac{dA}{dt}\Big).$$

Example 13.41. Let $A(t) = \begin{bmatrix} f(t) & g(t) \\ h(t) & k(t) \end{bmatrix}$. Then

$$\begin{aligned}\frac{d}{dt}\det(A) &= \mathrm{tr}\Big(\begin{bmatrix} k & -g \\ -h & f \end{bmatrix}\begin{bmatrix} f' & g' \\ h' & k' \end{bmatrix}\Big) = \mathrm{tr}\Big(\begin{bmatrix} kf' - gh' & kg' - gk' \\ -hf' + fh' & -hg' + fk' \end{bmatrix}\Big) \\ &= kf' - gh' - hg' + fk' = (fk - gh)'.\end{aligned}$$

• Taylor's theorem for functions of two variables

Let $R \subseteq \mathbb{R}^2$ be an open region, and $f : R \to \mathbb{R}$ be a function of two variables. For $(x, y), (a, b) \in R$, define $g(t) = f\big(a + t(x-a), b + t(y-b)\big)$. Suppose that all the k-th partial derivatives of f are continuous for $0 \leqslant k \leqslant n+1$ (which, by Theorem 13.35, implies that g is $(n+1)$-times differentiable), then Taylor's Theorem implies that there exists $\xi \in (0, 1)$ such that

$$g(1) = \sum_{k=0}^{n} \frac{g^{(k)}(0)}{k!} + \frac{g^{(n+1)}(\xi)}{(n+1)!}.$$

Now we compute $g^{(k)}(0)$. First by the chain rule,

$$\begin{aligned} g'(t) &= \frac{d}{dt} f\big(a + t(x-a), b + t(y-b)\big) \\ &= f_x\big(a + t(x-a), b + t(y-b)\big)(x-a) + f_y\big(tx + (1-t)a, ty + (1-t)b\big)(y-b); \end{aligned}$$

thus $g'(0) = f_x(a, b)(x-a) + f_y(a, b)(y-b)$. In general, we can prove by induction that

$$g^{(k)}(t) = \sum_{j=0}^{k} C^k_j \frac{\partial^k f}{\partial x^{k-j}\partial y^j}\big(a + t(x-a), b + t(y-b)\big)(x-a)^{k-j}(y-b)^j \tag{13.5.5}$$

under the assumption that the k-th partial derivatives are continuous (on an open region containing the line segment connecting (x, y) and (a, b)). To see this, we first simplify the notation by letting $\gamma(t) = \big(a + t(x-a), b + t(y-b)\big)$. We note that (13.5.5) holds for $k = 1$. Suppose that (13.5.5) holds for $k = \ell$. Then by the chain rule and Theorem 13.28, we find that

$$\begin{aligned}
g^{(\ell+1)}(t) &= \frac{d}{dt} g^{(\ell)}(t) = \frac{d}{dt} \sum_{j=0}^{\ell} C^{\ell}_{j} \frac{\partial^{\ell} f}{\partial x^{\ell-j} \partial y^{j}} \big(\gamma(t)\big)(x-a)^{\ell-j}(y-b)^{j} \\
&= \sum_{j=0}^{\ell} C^{\ell}_{j} \Big[\frac{\partial^{\ell+1} f}{\partial x^{\ell-j+1} \partial y^{j}} \big(\gamma(t)\big)(x-a)^{\ell-j+1}(y-b)^{j} \\
&\qquad\qquad + \frac{\partial^{\ell+1} f}{\partial x^{\ell-j} \partial y^{j+1}} \big(\gamma(t)\big)(x-a)^{\ell-j}(y-b)^{j+1} \Big] \\
&= \sum_{j=0}^{\ell} C^{\ell}_{j} \frac{\partial^{\ell+1} f}{\partial x^{\ell+1-j} \partial y^{j}} \big(\gamma(t)\big)(x-a)^{\ell+1-j}(y-b)^{j} \\
&\quad + \sum_{j=1}^{\ell+1} C^{\ell}_{j-1} \frac{\partial^{\ell+1} f}{\partial x^{\ell+1-j} \partial y^{j}} \big(\gamma(t)\big)(x-a)^{\ell+1-j}(y-b)^{j} \\
&= \frac{\partial^{\ell+1} f}{\partial x^{\ell+1}} \big(\gamma(t)\big)(x-a)^{\ell+1} + \frac{\partial^{\ell+1} f}{\partial y^{\ell+1}} \big(\gamma(t)\big)(y-b)^{\ell+1} \\
&\quad + \sum_{j=1}^{\ell} (C^{\ell}_{j} + C^{\ell}_{j-1}) \frac{\partial^{\ell+1} f}{\partial x^{\ell+1-j} \partial y^{j}} \big(\gamma(t)\big)(x-a)^{\ell+1-j}(y-b)^{j} \,.
\end{aligned}$$

By Pascal's Theorem,

$$\begin{aligned}
g^{(\ell+1)}(t) &= \frac{\partial^{\ell+1} f}{\partial x^{\ell+1}} \big(\gamma(t)\big)(x-a)^{\ell+1} + \frac{\partial^{\ell+1} f}{\partial y^{\ell+1}} \big(\gamma(t)\big)(y-b)^{\ell+1} \\
&\quad + \sum_{j=1}^{\ell} C^{\ell+1}_{j} \frac{\partial^{\ell+1} f}{\partial x^{\ell+1-j} \partial y^{j}} \big(\gamma(t)\big)(x-a)^{\ell+1-j}(y-b)^{j} \\
&= \sum_{j=0}^{\ell+1} C^{\ell+1}_{j} \frac{\partial^{\ell+1} f}{\partial x^{\ell+1-j} \partial y^{j}} \big(\gamma(t)\big)(x-a)^{\ell+1-j}(y-b)^{j} \,;
\end{aligned}$$

thus we establish (13.5.5) by induction. Therefore, by the fact that $g(1) = f(x, y)$ and $g(0) = f(a, b)$,

$$f(x,y) = \sum_{k=0}^{n} \frac{1}{k!} \sum_{j=0}^{k} C^{k}_{j} \frac{\partial^{k} f}{\partial x^{k-j} \partial y^{j}} (a,b)(x-a)^{k-j}(y-b)^{j} + R_n(x,y) \,, \tag{13.5.6}$$

where

$$R_n(x,y) = \frac{1}{(n+1)!}\sum_{j=0}^{n+1} C_j^{n+1}\frac{\partial^k f}{\partial x^{k-j}\partial y^j}\big(a+\xi(x-a), b+\xi(y-b)\big)(x-a)^{n+1-j}(y-b)^j\,.$$

The "polynomial" of two variables

$$P_n(x,y) = \sum_{k=0}^{n}\frac{1}{k!}\sum_{j=0}^{k} C_j^k \frac{\partial^k f}{\partial x^{k-j}\partial y^j}(a,b)(x-a)^{k-j}(y-b)^j$$

is called the n-th Taylor polynomial for f centered at (a,b), and the function R_n is the remainder associated with P_n.

Expanding the sum, we find that

$$\begin{aligned}
P_n(x,y) = {} & f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)\\
& + \frac{1}{2!}\Big[f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2\Big]\\
& + \frac{1}{3!}\Big[f_{xxx}(a,b)(x-a)^3 + 3f_{xxy}(a,b)(x-a)^2(y-b) + 3f_{xyy}(x-a)(y-b)^2\\
& \quad + f_{yyy}(a,b)(y-b)^3\Big] + \cdots +\\
& + \frac{1}{n!}\Big[\frac{\partial^n f}{\partial x^n}(a,b)(x-a)^n + C_1^n\frac{\partial^n f}{\partial x^{n-1}\partial y}(a,b)(x-a)^{n-1}(y-b) + \cdots\cdots +\\
& \quad + C_{n-1}^n\frac{\partial^n f}{\partial x\partial y^{n-1}}(a,b)(x-a)(y-b)^{n-1} + \frac{\partial^n f}{\partial y^n}(a,b)(y-b)^n\Big]\,.
\end{aligned}$$

Example 13.42. Find the third Taylor polynomial for the function $f(x,y) = \sin(xy)$ centered at $(0,0)$.

We compute the first, the second and the third partial derivatives of f as follows:

$$\begin{aligned}
f_x(x,y) &= y\cos(xy)\,, \quad f_y(x,y) = x\cos(xy)\,,\\
f_{xx}(x,y) &= -y^2\sin(xy)\,, \quad f_{xy}(x,y) = \cos(xy) - xy\sin(xy)\,, \quad f_{yy}(x,y) = -x^2\sin(xy)\,,\\
f_{xxx}(x,y) &= -y^3\cos(xy)\,, \quad f_{xxy}(x,y) = -2y\sin(xy) - xy^2\cos(xy)\,,\\
f_{xyy}(x,y) &= -2x\sin(xy) - x^2y\cos(xy)\,, \quad f_{yyy}(x,y) = -x^3\cos(xy)\,.
\end{aligned}$$

Therefore, the only non-vanishing term, when plugging $(x,y) = (0,0)$, is $f_{xy}(0,0) = 1$; thus

$$P_3(x,y) = \frac{1}{2!}\cdot 2f_{xy}(0,0)(x-0)(y-0) = xy\,.$$

Example 13.43. Find the second Taylor polynomial for the function $f(x, y) = \exp(x^2 + 2y)$ centered at $(0, 0)$.

We compute the first and the second partial derivatives of f as follows:

$$\begin{aligned} f_x(x, y) &= 2x \exp(x^2 + 2y)\,, \quad f_y(x, y) = 2 \exp(x^2 + 2y)\,, \\ f_{xx}(x, y) &= (2 + 4x^2) \exp(x^2 + 2y)\,, \quad f_{xy}(x, y) = 4x \exp(x^2 + 2y)\,, \\ f_{yy}(x, y) &= 4 \exp(x^2 + 2y)\,. \end{aligned}$$

Therefore, $f_x(0, 0) = f_{xy}(0, 0) = 0$, $f_y(0, 0) = f_{xx}(0, 0) = 2$, $f_{yy}(0, 0) = 4$; thus

$$\begin{aligned} P_2(x, y) &= f(0, 0) + f_x(0, 0)x + f_y(0, 0)y + \frac{1}{2!}\big[f_{xx}(0, 0)x^2 + 2f_{xy}(0, 0)xy + f_{yy}(0, 0)y^2\big] \\ &= 1 + 2y + x^2 + 2y^2\,. \end{aligned}$$

• **Implicit partial differentiation**

In Section 2.4 we have talked about finding derivatives of a function $y = f(x)$ which is defined implicitly by $F(x, y) = 0$ (when F is giving explicitly). Now suppose that $z = F(x, y)$ is a differentiable function and the relation $F(x, y) = 0$ defines a differentiable function $y = f(x)$ implicitly (so that $F(x, f(x)) = 0$). By the chain rule,

$$0 = \frac{d}{dx}F(x, f(x)) = F_x(x, f(x)) + F_y(x, f(x))f'(x)$$

which implies that

$$f'(x) = -\frac{F_x(x, f(x))}{F_y(x, f(x))} \quad \text{if} \quad F_y(x, f(x)) \neq 0\,.$$

Since f is in general unknown (but exists), we usually write the identity above as

$$\frac{dy}{dx} = -\frac{F_x(x, y)}{F_y(x, y)} \quad \text{if } F(x, y) = 0 \text{ and } F_y(x, y) \neq 0\,.$$

In fact, when F_x and F_y are continuous in an open region R, and $F(a, b) = 0$ and $F_y(a, b) \neq 0$ at some point $(a, b) \in R$, the relation $F(x, y) = 0$ defines a function $y = f(x)$ implicitly near (a, b) and f is continuously differentiable near $x = a$. This is the Implicit Function Theorem and the precise statement is stated as follows.

Theorem 13.44: Implicit Function Theorem (Special case)

Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $F : R \to \mathbb{R}$ be a function of two variables such that F_x and F_y are continuous in a neighborhood of $(a, b) \in R$. If $F(a, b) = 0$ and $F_y(a, b) = 0$, then there exists $\delta > 0$ and a unique function $f : (a - \delta, a + \delta) \to \mathbb{R}$ satisfying $F(x, f(x)) = 0$ for all $x \in (a - \delta, a + \delta)$, and $b = f(a)$. Moreover, f is differentiable on $(a - \delta, a + \delta)$, and

$$f'(x) = -\frac{F_x(x, f(x))}{F_y(x, f(x))} \qquad \forall\, x \in (a - \delta, a + \delta)\,.$$

In general, if F is a function of n variables $(x_1, x_2, \cdots, x_n)$ such that F_{x_1}, F_{x_2}, $\cdots$, F_{x_n} are continuous in a neighborhood of $(a_1, a_2, \cdots, a_n$. If $F(a_1, a_2, \cdots, a_n) = 0$ and $F_{x_n}(a_1, a_2, \cdots, a_n) \neq 0$, then locally there exists a unique function f satisfying $F(x_1, \cdots, x_{n-1}, f(x_1, \cdots, x_{n-1})) = 0$ and $a_n = f(a_1, \cdots, a_{n-1})$. Moreover, for $1 \leqslant j \leqslant n - 1$,

$$\frac{\partial f}{\partial x_j}(x_1, \cdots, x_{n-1}) = -\frac{F_{x_j}(x_1, \cdots, x_{n-1}, f(x_1, \cdots, x_{n-1}))}{F_{x_n}(x_1, \cdots, x_{n-1}, f(x_1, \cdots, x_{n-1}))}\,.$$

Example 13.45. Find $\dfrac{dy}{dx}$ if (x, y) satisfies $y^3 + y^2 - 5y - x^2 + 4 = 0$.

Let $F(x, y) = y^3 + y^2 - 5y - x^2 + 4$. Then $F_x(x, y) = -2x$ and $F_y(x, y) = 3y^2 + 2y - 5$. Therefore,

$$\frac{dy}{dx} = -\frac{F_x(x, y)}{F_y(x, y)} = \frac{2x}{3y^2 + 2y - 5}\,.$$

Example 13.46. Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{\partial z}{\partial y}$ if (x, y, z) satisfies $3x^2z - x^2y^2 + 2z^3 + 3yz - 5 = 0$.

Let $F(x, y, z) = 3x^2z - x^2y^2 + 2z^3 + 3yz - 5$. Then $F_x(x, y, z) = 6xz - 2xy^2$, $F_y(x, y, z) = -2x^2y + 3z$ and $F_z(x, y, z) = 3x^2 + 6z^2 + 3y$. Therefore,

$$\frac{\partial z}{\partial x} = -\frac{F_x(x, y, z)}{F_z(x, y, z)} = \frac{2xy^2 - 6xz}{3x^2 + 6z^2 + 3y}$$

and

$$\frac{\partial z}{\partial y} = -\frac{F_y(x, y, z)}{F_z(x, y, z)} = \frac{2x^2y - 3z}{3x^2 + 6z^2 + 3y}\,.$$

13.6 Directional Derivatives and Gradients

Let f be a function of two variables. From the discussion above we know that the existence of f_x and f_y does not guarantee the differentiability of f. Since f_x and f_y are the rate of change of the function f in two special directions $(1,0)$ and $(0,1)$, we can ask ourselves whether f is differentiable if the rate of change of f exist in all direction.

Definition 13.47

Let f be a function of two variables x and y, and let $\boldsymbol{u} = \cos\theta\mathbf{i}+\sin\theta\mathbf{j}$, where $\mathbf{i} = (1,0)$ and $\mathbf{j} = (0,1)$, be a unit vector. The directional derivative of f in the direction of $\boldsymbol{u}$ at (a,b), denoted by $D_{\boldsymbol{u}}f(a,b)$, is the limit

$$D_{\boldsymbol{u}}f(a,b) = \lim_{h\to 0}\frac{f(a+h\cos\theta, b+h\sin\theta) - f(a,b)}{h}$$

provided this limit exists.

Example 13.48. Find the direction derivative of $f(x,y) = x^2\sin 2y$ at $\big(1,\frac{\pi}{2}\big)$ in the direction of $\boldsymbol{v} = 3\mathbf{i} - 4\mathbf{j}$.

We first normalize the vector $\boldsymbol{v}$ and find that $\boldsymbol{u} = \frac{3}{5}\mathbf{i} - \frac{4}{5}\mathbf{j}$ is in the same direction of $\boldsymbol{v}$ and has unit length. Therefore, for $h \neq 0$,

$$\frac{f\big(1+\frac{3h}{5}, \frac{\pi}{2} - \frac{4h}{5}\big) - f\big(1,\frac{\pi}{2}\big)}{h} = \frac{(1+\frac{3h}{5})^2\sin\big(\pi - \frac{8h}{5}\big) - 1^2\sin\pi}{h} = \Big(1+\frac{3h}{5}\Big)^2\frac{\sin\frac{8h}{5}}{h};$$

thus by the fact that $\lim\limits_{h\to 0}\frac{\sin h}{h} = 1$, we find that

$$\lim_{h\to 0}\frac{f\big(1+\frac{3h}{5}, \frac{\pi}{2} - \frac{4h}{5}\big) - f\big(1,\frac{\pi}{2}\big)}{h} = \lim_{h\to 0}\Big(1+\frac{3h}{5}\Big)^2\frac{\sin\frac{8h}{5}}{h} = \frac{8}{5}.$$

Theorem 13.49

Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $f : R \to \mathbb{R}$ be a function of two variables. If f is differentiable at $(x_0,y_0) \in R$, then for all unit vector $\boldsymbol{v} = \cos\theta\mathbf{i}+\sin\theta\mathbf{j}$,

$$(D_{\boldsymbol{u}}f)(x_0,y_0) = f_x(x_0,y_0)\cos\theta + f_y(x_0,y_0)\sin\theta = (Df)(x_0,y_0)\cdot\boldsymbol{u}.$$

Proof. Let $g(t) = f(x_0 + t\cos\theta, y_0 + t\sin\theta)$. Then by the chain rule for functions of two variables,

$$(D_{\boldsymbol{u}}f)(x_0, y_0) = \lim_{h\to 0} \frac{g(h) - g(0)}{h} = g'(0) = f_x(x_0, y_0)\cos\theta + f_y(x_0, y_0)\sin\theta\,. \qquad \square$$

Example 13.50. In this example we re-compute of the direction derivative in Example 13.48 using Theorem 13.49. Note that $f(x, y) = x^2 \sin 2y$ is differentiable on $\mathbb{R}^2$ since $f_x(x, y) = 2x \sin 2y$ and $f_y(x, y) = 2x^2 \cos 2y$ are continuous (so that Theorem 13.35) guarantees the differentiability of f). Therefore, Theorem 13.49 implies that

$$(D_{\boldsymbol{u}}f)\big(1, \frac{\pi}{2}\big) = \frac{3}{5} f_x\big(1, \frac{\pi}{2}\big) - \frac{4}{5} f_y\big(1, \frac{\pi}{2}\big) = \frac{3}{5} \cdot 2 \cdot \sin\pi - \frac{4}{5} \cdot 2 \cdot 1^2 \cdot \cos\pi = \frac{8}{5}\,.$$

Unfortunately, the existence of directional derivative of f in all directions does not imply the differentiability of f.

Example 13.51. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

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

and $\boldsymbol{u} = (\cos\theta, \sin\theta) \in \mathbb{R}^2$ be a unit vector. Then if $\cos\theta \neq 0$ $\big($or equivalently, $\theta \neq \frac{\pi}{2}, \frac{3\pi}{2}\big)$,

$$(D_{\boldsymbol{u}}f)(0, 0) = \lim_{h\to 0} \frac{f(h\cos\theta, h\sin\theta) - f(0, 0)}{h} = \lim_{t\to 0} \frac{h^3 \cos\theta \sin\theta^2}{h(h^2\cos\theta^2 + h^4 \sin\theta^4)} = \frac{\sin\theta^2}{\cos\theta}$$

while if $\cos\theta = 0$,

$$(D_{\boldsymbol{u}}f)(0, 0) = \lim_{h\to 0} \frac{f(h\cos\theta, h\sin\theta) - f(0, 0)}{h} = 0\,.$$

Therefore, the directional derivative of f at $(0, 0)$ exist in all directions. However, f is not continuous at $(0, 0)$ since if (x, y) approaches $(0, 0)$ along the curve $x = my^2$ with $m \neq 0$, we have

$$\lim_{\substack{(x,y)\to(0,0) \\ x=my^2}} f(x, y) = \lim_{y\to 0} f(my^2, y) = \lim_{y\to 0} \frac{my^4}{m^2y^4 + y^4} = \frac{m}{m^2 + 1}$$

which depends on m. Therefore, f is not continuous at $(0, 0)$.

Definition 13.52

Let $z = f(x, y)$ be a function of x and y such that $f_x(a, b)$ and $f_y(a, b)$ exists. Then the gradient of f at (a, b), denoted by $(\nabla f)(a, b)$ or $(\textbf{grad}f)(a, b)$, is the vector $(f_x(a, b), f_y(a, b))$; that is,

$$(\nabla f)(a, b) = \big(f_x(a, b), f_y(a, b)\big) = f_x(a, b)\mathbf{i} + f_y(a, b)\mathbf{j}\,.$$

- **Functions of several variables**

Definition 13.53

Let f be a function of n variables. The directional derivative of f at $(a_1, a_2, \cdots, a_n)$ in the direction $\boldsymbol{u} = (u_1, u_2, \cdots, u_n)$, where $u_1^2 + u_2^2 + \cdots + u_n^2 = 1$, is the limit

$$(D_{\boldsymbol{u}} f)(a_1, a_2, \cdots, a_n) = \lim_{h \to 0} \frac{f(a_1 + hu_1, a_2 + hu_2, \cdots, a_n + hu_n) - f(a_1, a_2, \cdots, a_n)}{h}$$

provided that the limit exists. The gradient of f at $(a_1, a_2, \cdots, a_n)$, denoted by $(\nabla f)(a_1, a_2, \cdots, a_n)$, is the vector

$$(\nabla f)(a_1, a_2, \cdots, a_n) = \big(f_{x_1}(a_1, \cdots, a_n), f_{x_2}(a_1, \cdots, a_n), \cdots, f_{x_n}(a_1, \cdots, a_n)\big) .$$

Theorem 13.54

Let f be a function of n variables. If f is differentiable at $(a_1, a_2, \cdots, a_n)$ and $\boldsymbol{u} = (u_1, u_2, \cdots, u_n)$ is a unit vector, then

$$(D_{\boldsymbol{u}} f)(a_1, a_2, \cdots, a_n) = (\nabla f)(a_1, \cdots, a_n) \cdot \boldsymbol{u} .$$

- **Properties of the gradient**

Theorem 13.55

Let f be a function of two variables. If f has continuous first partial derivatives f_x and f_y in a neighborhood of (x_0, y_0) and $(\nabla f)(x_0, y_0) \neq \mathbf{0}$, then $(\nabla f)(x_0, y_0)$ is perpendicular/normal to the level curve $f(x, y) = f(x_0, y_0)$ at (x_0, y_0). Moreover, the value of f at (x_0, y_0) increase most rapidly in the direction $\dfrac{(\nabla f)(x_0, y_0)}{\|(\nabla f)(x_0, y_0)\|}$ and decreases most rapidly in the direction $-\dfrac{(\nabla f)(x_0, y_0)}{\|(\nabla f)(x_0, y_0)\|}$, where $\|\cdot\|$ denotes the length of the vector.

Remark 13.56. 1. Let $f : (a, b) \to \mathbb{R}$ be differentiable. The graph of the function $y = f(x)$ can be view as the level set $F(x, y) = y - f(x)$ through point $(c, f(c))$ (that is, $F(x, y) = F(c, f(c))$). We note that at the slope of the tangent line $(c, f(c))$ if $f'(c)$ (so that $(1, f'(c))$ is a tangent vector at $(c, f(c))$); thus the vector $(-f'(c), 1)$ is perpendicular to the graph of f at $(c, f(c))$. The theorem above generalizes this result.

2. The terminology "the value of f at (x_0, y_0) increase most rapidly in the direction $\boldsymbol{u}$", where $\boldsymbol{u}$ is a unit vector, means that the directional derivative $(D_{\boldsymbol{v}}f)(x_0, y_0)$, treated as a function of $\boldsymbol{v}$, attains its maximum at $\boldsymbol{v} = \boldsymbol{u}$.

Example 13.57. Let $f(x, y) = \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$. Then the level curve $f(x, y) = 1$ is an ellipse and the normal vector of this level curve at point $(a\cos\theta, b\sin\theta)$ is given by

$$\big(f_x(a\cos\theta, b\sin\theta), f_y(a\cos\theta, b\sin\theta)\big) = \big(\frac{2\cos\theta}{a}, \frac{2\sin\theta}{b}\big).$$

Example 13.58. A heat-seeking particle is located at the point $(2, -3)$ on a metal plate whose temperature at (x, y) is $T(x, y) = 20 - 4x^2 - y^2$. Find the path of the particle as it continuously moves in the direction of maximum temperature increase.

Suppose the path of the particle is given by $(x(t), y(t))$. Then

$$(x'(t), y'(t)) /\!/ (\nabla T)(x(t), y(t)) = \big(-8x(t), -2y(t)\big).$$

Therefore, there exists a function $k(t)$ such that $-8x = k\dfrac{dx}{dt}$ and $-2y = k\dfrac{dy}{dt}$; thus

$$\frac{d}{dt}\big(\ln|x| - 4\ln|y|\big) = 0.$$

Then $|x||y|^{-4} = C$. Since $(x(t), y(t))$ passes through $(2, -3)$, we find that $C = \dfrac{2}{81}$; thus (x, y) satisfies $x = \dfrac{2}{81}y^4$.

Theorem 13.59

Let f be a function of three variables. If f has continuous first partial derivatives f_x, f_y, f_z in a neighborhood of (x_0, y_0, z_0) and $(\nabla f)(x_0, y_0, z_0) \neq \mathbf{0}$, then $(\nabla f)(x_0, y_0, z_0)$ is perpendicular/normal to the level surface $f(x, y, z) = f(x_0, y_0, z_0)$ at (x_0, y_0, z_0). Moreover, the value of f at (x_0, y_0, z_0) increase most rapidly in the direction $\dfrac{(\nabla f)(x_0, y_0, z_0)}{\|(\nabla f)(x_0, y_0, z_0)\|}$ and decreases most rapidly in the direction $-\dfrac{(\nabla f)(x_0, y_0, z_0)}{\|(\nabla f)(x_0, y_0, z_0)\|}$, where $\|\cdot\|$ denotes the length of the vector.

Proof. We have shown that $(\nabla F)(x_0, y_0, z_0)$ is perpendicular to the level surface $F(x, y, z) = F(x_0, y_0, z_0)$ in Theorem 13.63, so it suffices to show that $(D_{\boldsymbol{v}}F)(x_0, y_0, z_0)$ attains its maximum at $\boldsymbol{v} = \boldsymbol{u}$. Nevertheless, by Theorem 13.54, we find that

$$(D_{\boldsymbol{v}}F)(x_0, y_0, z_0) = (\nabla F)(x_0, y_0, z_0) \cdot \boldsymbol{v} = \|(\nabla F)(x_0, y_0, z_0)\| \cos\theta,$$

where θ is the angle between $(\nabla F)(x_0, y_0, z_0)$ and $\boldsymbol{v}$. Clearly $(D_{\boldsymbol{v}}F)(x_0, y_0, z_0)$ attains its maximum when $\theta = 0$ which shows that $(D_{\boldsymbol{v}}F)(x_0, y_0, z_0)$ attains its maximum at $\boldsymbol{v} = \dfrac{(\nabla F)(x_0, y_0, z_0)}{\|(\nabla F)(x_0, y_0, z_0)\|}$. □

Example 13.60 (Gradient method of finding local minimum of a function)**.** Suppose that you are looking for the minimum of a function $f : \mathbb{R}^2 \to \mathbb{R}$. You do not know where the minimum point of f is, so you start with (conjecturing a possible) point (a, b) and hope to find a curve C that connects (a, b) and the minimum point. Suppose that C is parameterized by $\boldsymbol{r} : [a, b] \to \mathbb{R}^2$. By the fact that $-(\nabla f)(\boldsymbol{x})$ points to the direction to which f decreases most rapidly, we expected that

$$\boldsymbol{r}'(t) \,/\!/\, -(\nabla f)(\boldsymbol{r}(t)) \,.$$

In particular, we choose $\boldsymbol{r}'(t) = -(\nabla f)(\boldsymbol{r}(t))$ and hope that we can find $\boldsymbol{r}$ (so that we can find C). We note that we can also choose $\boldsymbol{r}'(t) = -\dfrac{(\nabla f)(\boldsymbol{r}(t))}{\|(\nabla f)(\boldsymbol{r}(t))\|}$ which implies that $\boldsymbol{r}'$ never vanishes so that the tangent direction indeed points to the direction $-(\nabla f)(\boldsymbol{r}(t))$.

Sometimes it is very hard to find the solution $\boldsymbol{r}$ to the differential equationt, so instead we choose a different strategy. Starting at the point (a, b), we move forward in the direction $-(\nabla f)(a, b)$ and stop temporally at $(a_1, b_1) \equiv (a, b) - t_0(\nabla f)(a, b)$ for some $t > 0$. Then we move forward in the direction $-(\nabla f)(a_1, b_1)$ and stop temporally at $(a_2, b_2) \equiv (a_1, b_1) - t_1(\nabla)(a_1, b_1)$. Continue this process, we obtain a sequence of stops $\{(a_k, b_k)\}_{k=1}^{\infty}$ given by

$$(a_{k+1}, b_{k+1}) = (a_k, b_k) - t_k(\nabla f)(a_k, b_k) \tag{13.6.1}$$

for some sequence $\{t_k\}_{k=0}^{\infty}$ of non-negative numbers to be chosen. One way of choosing the step-size t_k, called the method of exact line search, is to choose t_k so that

$$f\big((a_k, b_k) - t_k(\nabla f)(a_k, b_k)\big) = \min_{t>0} f\big((a_k, b_k) - t(\nabla f)(a_k, b_k)\big) \,.$$

Such t_k must satisfy that

$$\frac{d}{dt}\Big|_{t=t_k} f\big((a_k, b_k) - t(\nabla f)(a_k, b_k)\big) = 0$$

which implies that t_k satisfies that $(\nabla f)\big((a_k, b_k) - t_k(\nabla f)(a_k, b_k)\big) \cdot (\nabla f)(a_k, b_k) = 0$. Therefore, (13.6.1) implies that

$$(\nabla f)(a_{k+1}, b_{k+1}) \cdot (\nabla f)(a_k, b_k) = 0 \qquad \forall\, k \in \mathbb{N} \cup \{0\}$$

which shows that the exact line search algorithm of constructing minimizing sequence produces a zigzag path connecting the starting point and the minimum point.

13.7 Tangent Planes and Normal Lines

• The tangent plane of surfaces

Any three points in space that are not collinear defines a plane. Suppose that $\mathcal{S}$ is a "surface" (which we have not define yet, but please use the common sense to think about it), and $P_0 = (x_0, y_0, z_0)$ is a point on the plane. Given another two point $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ on the surface such that P_0, P_1, P_2 are not collinear, let $T_{P_1P_2}$ denote the plane determined by P_0, P_1 and P_2. If the plane "approaches" a certain plane as P_1, P_2 approaches P_0, the "limit" is called the tangent plane of $\mathcal{S}$ at P_0.

Now suppose that the surface $\mathcal{S}$ is the graph of a function of two variables $z = f(x, y)$. Consider the tangent plane of $\mathcal{S}$ at $P_0 = (x_0, y_0, z_0)$, where $z_0 = f(x_0, y_0)$. The plane $T_{P_1P_2}$, where $P_1 = (x_0 + h, y_0, f(x_0 + h, y_0))$ and $P_2 = (x_0, y_0 + k, f(x_0, y_0 + k))$, is given by

$$\Big[\big(h, 0, f(x_0+h, y_0) - f(x_0, y_0)\big) \times \big(0, k, f(x_0, y_0+k) - f(x_0, y_0)\big)\Big] \cdot (x - x_0, y - y_0, z - z_0) = 0\,,$$

where $\boldsymbol{u} \cdot \boldsymbol{v}$ and $\boldsymbol{u} \times \boldsymbol{v}$ are the inner product and the cross product of $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively. For $(h, k) \neq (0, 0)$, divide both sides by hk and pass to the limit as $(h, k) \to (0, 0)$, we find that the limit is

$$\Big[\big(1, 0, f_x(x_0, y_0)\big) \times \big(0, 1, f_y(x_0, y_0)\big)\Big] \cdot (x - x_0, y - y_0, z - z_0) = 0\,,$$

provided that $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ exists. Computing the cross product, we find that

$$\big(1, 0, f_x(x_0, y_0)\big) \times \big(0, 1, f_y(x_0, y_0)\big) = \big(-f_x(x_0, y_0), -f_y(x_0, y_0), 1\big)\,;$$

thus if the tangent plane exists at (x_0, y_0, z_0), the tangent plane must be

$$\big(-f_x(x_0, y_0), -f_y(x_0, y_0), 1\big) \cdot \big(x - x_0, y - y_0, z - f(x_0, y_0)\big) = 0$$

or equivalently,

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)\,.$$

On the other hand, if f is differentiable at (x_0, y_0), then

$$\begin{aligned} f(x, y) = f(x_0, y_0) &+ f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) \\ &+ \varepsilon_1(x, y)(x - x_0) + \varepsilon_2(x, y)(y - y_0) \end{aligned}$$

for some functions ε_1, ε_2 satisfying $\lim\limits_{(x,y)\to(x_0,y_0)} \varepsilon_1(x,y) = \lim\limits_{(x,y)\to(x_0,y_0)} \varepsilon_2(x,y) = 0$. This shows that the rate of convergence of the quantity

$$\big|f(x,y) - f(x_0,y_0) - f_x(x_0,y_0)(x-x_0) - f_y(x_0,y_0)(y-y_0)\big|\,,$$

as (x,y) approaches (x_0,y_0), is "faster than linear" and this is exactly what we have in mind when talking about tangent planes. Therefore, we conclude that

Theorem 13.61

Let $R \subseteq \mathbb{R}^2$ be an open region in the plane, and $f : R \to \mathbb{R}$ be a function of two variables. If f is differentiable at $(x_0,y_0) \in R$, the tangent plane of the graph of f at $(x_0,y_0,f(x_0,y_0))$ is given by

$$z = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)\,,$$

and the vector $(f_x(x_0,y_0), f_y(x_0,y_0), -1)$ is a normal vector to the graph of f at $(x_0,y_0,f(x_0,y_0))$.

Example 13.62. Find the equation of the normal line to the surface $xyz = 12$ at the point $(2,-2,-3)$.

Let $F(x,y,z) = xyz - 12$. Then $(F_x,F_y,F_z)(2,-2,-3) = (6,-6,-4)$. Therefore, the vector $(6,-6,-4)$ is normal to the surface $xyz = 12$ at $(2,-2,-3)$ and the normal line passing through $(2,-2,-3)$ is

$$\frac{x-2}{6} = \frac{y+2}{-6} = \frac{z+3}{-4}\,.$$

Now suppose that the function of three variables $w = F(x,y,z)$ is continuously differentiable; that is, F_x, F_y, F_z are continuous. Suppose that for some (x_0,y_0,z_0) in the domain, $(F_x(x_0,y_0,z_0), F_y(x_0,y_0,z_0), F_z(x_0,y_0,z_0)) \neq \mathbf{0}$. W.L.O.G., we assume that $F_z(x_0,y_0,z_0) \neq 0$. Then the Implicit Function Theorem (Theorem 13.44) implies that there exists a unique differentiable function $z = f(x,y)$ such that

$$F(x,y,f(x,y)) = 0 \qquad \text{and} \qquad z_0 = f(x_0,y_0)\,.$$

By the discussion above, the tangent plane of the graph of f at (x_0,y_0,z_0) is given by

$$z = z_0 + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

and the implicit partial differentiation further shows that the tangent plane above can be rewritten as

$$z = z_0 - \frac{F_x(x_0, y_0, z_0)}{F_z(x_0, y_0, z_0)}(x - x_0) - \frac{F_y(x_0, y_0, z_0)}{F_z(x_0, y_0, z_0)}(y - y_0)\,.$$

Therefore, the tangent plane of the graph of f at (x_0, y_0, z_0) is given by

$$\big(F_x(x_0, y_0, z_0), F_y(x_0, y_0, z_0), F_z(x_0, y_0, z_0)\big) \cdot (x - x_0, y - y_0, z - z_0) = 0\,.$$

On the other hand, note that the graph of f is the same as the level surface $F(x, y, z) = F(x_0, y_0, z_0)$; thus we conclude that

Theorem 13.63

Let $w = F(x, y, z)$ be a function of three variables such that F_x, F_y and F_z are continuous. If $\big(F_x(x_0, y_0, z_0), F_y(x_0, y_0, z_0), F_z(x_0, y_0, z_0)\big) \neq \mathbf{0}$, then the tangent plane of the level surface $F(x, y, z) = F(x_0, y_0, z_0)$ at (x_0, y_0, z_0) is given by

$$\big(F_x(x_0, y_0, z_0), F_y(x_0, y_0, z_0), F_z(x_0, y_0, z_0)\big) \cdot (x - x_0, y - y_0, z - z_0) = 0\,,$$

and the vector $\big(F_x(x_0, y_0, z_0), F_y(x_0, y_0, z_0), F_z(x_0, y_0, z_0)\big)$ is a normal vector to the level surface $F(x, y, z) = F(x_0, y_0, z_0)$.

Example 13.64. Find an equation of the normal line and the tangent plane to the paraboloid

$$z = 1 - \frac{1}{10}(x^2 + 4y^2)$$

at the point $\big(1, 1, \frac{1}{2}\big)$.

Let $F(x, y, z) = z - 1 + \frac{1}{10}(x^2 + 4y^2)$. Then $F_z\big(1, 1, \frac{1}{2}\big) \equiv \big(\frac{1}{5}, \frac{4}{5}, 1\big) \neq \mathbf{0}$; thus Theorem 13.63 implies that the tangent plane of the given paraboloid at $\big(1, 1, \frac{1}{2}\big)$ is

$$z = \frac{1}{2} - \frac{1}{5}(x - 1) - \frac{4}{5}(y - 1) = \frac{3}{2} - \frac{1}{5}x - \frac{4}{5}y\,.$$

An equation of the normal line at $\big(1, 1, \frac{1}{2}\big)$ is given by

$$\frac{x - 1}{1/5} = \frac{y - 1}{4/5} = \frac{z - 1/2}{1}\,.$$

13.8 Extrema of Functions of Several Variables

13.8.1 Absolute extrema and relative extrema

Theorem 13.65: Extreme Value Theorem

Let f be a continuous function of two variables x and y defined on a closed bounded region R in the plane.

1. There is at least one point in R at which f takes on a minimum value.
2. There is at least one point in R at which f takes on a maximum value.

A minimum is also called an absolute minimum and a maximum is also called an absolute maximum. As in the case of functions of one variable, there are relative extrema defined as follows.

Definition 13.66: Relative Extrema

Let f be a function defined on a region R containing (x_0, y_0).

1. The function f has a relative minimum at (x_0, y_0) if $f(x, y) \geqslant f(x_0, y_0)$ for all (x, y) in an open disk containing (x_0, y_0).
2. The function f has a relative maximum at (x_0, y_0) if $f(x, y) \leqslant f(x_0, y_0)$ for all (x, y) in an open disk containing (x_0, y_0).

Similar to the critical points for functions of one variable defined in Definition 3.4,we have the following

Definition 13.67: Critical Points

Let f be defined on an open region R containing (x_0, y_0). The point (x_0, y_0) is a critical point of f if one of the following is true.

1. $f_x(x_0, y_0) = 0$ and $f_y(x_0, y_0) = 0$;
2. $f_x(x_0, y_0)$ or $f_y(x_0, y_0)$ does not exist.

Similar to Theorem 3.5, we have the following necessary condition for points where f attains its relative extrema.

Theorem 13.68

Let R be an open region in the plane, and $f : R \to \mathbb{R}$ be continuous. If f has a relative extremum at (x_0, y_0) on an open region R, then (x_0, y_0) is a critical point of f.

Example 13.69. Determine the relative extrema of the function

$$f(x, y) = -x^3 + 4xy - 2y^2 + 1 .$$

First we find the critical points of f. Since f is differentiable, the critical points are those points at which the gradient of f is the zero vector. Since $f_x(x, y) = -3x^2 + 4y$ and $f_y(x, y) = 4x - 4y$, if (a, b) is a critical point of f, then $-3a^2 + 4b = 4a - 4b = 0$. Therefore, $(0, 0)$ and $\big(\frac{4}{3}, \frac{4}{3}\big)$ are the only critical points of f.

Note that $(0, 0)$ is not a relative extremum of f since $f(x, 0)$ does not attain its extremum at $x = 0$. Near $\big(\frac{4}{3}, \frac{4}{3}\big)$, we find that if $|h|, |k| \ll 1$,

$$\begin{aligned} f\Big(\frac{4}{3} + h, \frac{4}{3} + k\Big) &= -\Big(h + \frac{4}{3}\Big)^3 + 4\Big(\frac{4}{3} + h\Big)\Big(\frac{4}{3} + k\Big) - 2\Big(k + \frac{4}{3}\Big)^2 + 1 \\ &= -h^3 - 4h^2 - \frac{16h}{3} - \frac{64}{27} + 4\Big(\frac{16}{9} + \frac{4}{3}h + \frac{4}{3}k + hk\Big) - 2\Big(k^2 + \frac{8}{3}k + \frac{16}{9}\Big) + 1 \\ &= -h^3 - 4h^2 + 4hk - 2k^2 + f\Big(\frac{4}{3}, \frac{4}{3}\Big) \\ &= f\Big(\frac{4}{3}, \frac{4}{3}\Big) - 2(k - h)^2 - h^2(2 + h) \leqslant f\Big(\frac{4}{3}, \frac{4}{3}\Big) . \end{aligned}$$

Therefore, f has a relative maximum at $\big(\frac{4}{3}, \frac{4}{3}\big)$.

13.8.2 The second partials test

A critical point of a function of two variables do not always yield relative maxima or minima.

Definition 13.70

Let f be a function of two variables. A point (x_0, y_0) is a saddle point of f if (x_0, y_0) is a critical point of f but f does not attain its extrema at (x_0, y_0).

Theorem 13.71

Suppose that a function f of two variables has continuous second partial derivatives on an open region containing a point (a, b) for which $f_x(a, b) = f_y(a, b) = 0$. Let

$$D = f_{xx}(a, b)f_{yy}(a, b) - f_{xy}(a, b)^2 = \begin{vmatrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{yx}(a, b) & f_{yy}(a, b) \end{vmatrix}.$$

1. If $D > 0$ and $f_{xx}(a, b) > 0$, then f has a relative minimum at (a, b).
2. If $D > 0$ and $f_{xx}(a, b) < 0$, then f has a relative maximum at (a, b).
3. If $D < 0$, then $(a, b, f(a, b))$ is a saddle point.
4. The test is inconclusive if $D = 0$.

Example 13.72. Consider the relative extrema of the function given in Example 13.69. We have computed that $(0, 0)$ and $\left(\frac{4}{3}, \frac{4}{3}\right)$ are the only critical points of f.

1. The point $(0, 0)$: we compute the second partial derivatives and obtain that

$$f_{xx}(0, 0) = 0\,, \quad f_{xy}(0, 0) = 4 \quad \text{and} \quad f_{yy}(0, 0) = -4\,.$$

Therefore, $D = -16 < 0$ which implies that $(0, 0)$ is a saddle point.

2. The point $\left(\frac{4}{3}, \frac{4}{3}\right)$: we compute the second partial derivatives and obtain that

$$f_{xx}\left(\frac{4}{3}, \frac{4}{3}\right) = -8\,, \quad f_{xy}\left(\frac{4}{3}, \frac{4}{3}\right) = 4 \quad \text{and} \quad f_{yy}\left(\frac{4}{3}, \frac{4}{3}\right) = -4\,.$$

Therefore, $D = 16 > 0$. Since $f_{xx}\left(\frac{4}{3}, \frac{4}{3}\right) < 0$, f has a relative maximum at $\left(\frac{4}{3}, \frac{4}{3}\right)$.

Example 13.73. Find the absolute extrema of the function $f(x, y) = \sin(xy)$ on the closed region given by $0 \leqslant x \leqslant \pi$ and $0 \leqslant y \leqslant 1$.

From the partial derivatives

$$f_x(x, y) = y\cos(xy) \quad \text{and} \quad f_y(x, y) = x\cos(xy)\,,$$

we find that each point on the hyperbola $xy = \frac{\pi}{2}$ is a critical point of f. The value of f at each of these points is $\sin\frac{\pi}{2} = 1$ which is the maximum of the sine function. Therefore, the maximum of f is 1.

The minimum of f occurs at the boundary of the region.

1. $x = 0$ and $0 \leqslant y \leqslant 1$: then $f(x, y) = 0$.

2. $x = \pi$ and $0 \leqslant y \leqslant 1$: then $f(x, y) = \sin(\pi y)$. The critical points of the function $g(y) = \sin(\pi y)$ occurs at $y = \frac{1}{2}$ since $g'\big(\frac{1}{2}\big) = \pi \cos\big(\frac{\pi}{2}\big) = 0$. Since $g\big(\frac{1}{2}\big) = 1$ and $g(0) = g(1) = 0$, we find that the minimum of g is 0.

3. $y = 0$ and $0 \leqslant x \leqslant \pi$: then $f(x, y) = 0$.

4. $y = 1$ and $0 \leqslant x \leqslant \pi$: then $f(x, y) = \sin x$ whose minimum on $[0, \pi]$ is 0.

Therefore, the minimum of f is 0.

The concepts of relative extrema and critical points can be extended to functions of three or more variables. On the other hand, the second derivative test for functions of three or more variables are more tricky, and we will not talk about this until the course of Advance Calculus.

13.9 Applications of Extrema

Theorem 13.74

The least squares regression line for n points $\big\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\big\}$ is given by $y = ax + b$, where

$$a = \frac{n \sum\limits_{i=1}^{n} x_i y_i - \Big(\sum\limits_{i=1}^{n} x_i\Big)\Big(\sum\limits_{i=1}^{n} y_i\Big)}{n \sum\limits_{i=1}^{n} x_i^2 - \Big(\sum\limits_{i=1}^{n} x_i\Big)^2} \quad \text{and} \quad b = \frac{1}{n}\Big(\sum_{i=1}^{n} y_i - a \sum_{i=1}^{n} x_i\Big). \tag{13.9.1}$$

Proof. For $a, b \in \mathbb{R}$, define $S(a, b) = \sum\limits_{i=1}^{n} (ax_i + b - y_i)^2$. Then

$$\begin{aligned} \frac{\partial S}{\partial a}(a, b) &= 2 \sum_{i=1}^{n} (ax_i + b - y_i) x_i \,, \\ \frac{\partial S}{\partial b}(a, b) &= 2 \sum_{i=1}^{n} (ax_i + b - y_i) \,. \end{aligned}$$

The critical points (a, b) of S satisfies

$$a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i\,, \tag{13.9.2a}$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} 1 = \sum_{i=1}^{n} y_i \tag{13.9.2b}$$

which implies that (a, b) are given by (13.9.1). Clearly such (a, b) minimizes S. □

Remark 13.75. An easy way to memorize the equations (a, b) satisfies is given in this remark. We assume (even though in general it is a false assumption) that the line $y = ax+b$ passes through (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n). Then $y_i = ax_i + b$ for all $1 \leqslant i \leqslant n$; thus in matrix form, we have

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 & 1 \\ y_2 & 1 \\ \vdots & \vdots \\ y_n & 1 \end{bmatrix}.$$

Therefore,

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 & 1 \\ y_2 & 1 \\ \vdots & \vdots \\ y_n & 1 \end{bmatrix}$$

which implies (13.9.2).

13.10 Lagrange Multipliers

The concept of this section is to find the extrema of a function of several variables subject to certain constraints:

> Find extrema of the function $w = f(x_1, x_2, \cdots, x_n)$ when $(x_1, x_2, \cdots, x_n)$ satisfies $g_1(x_1, \cdots, x_n) = g_2(x_1, \cdots, x_n) = \cdots = g_m(x_1, \cdots, x_n) = 0$.

Theorem 13.76: Lagrange Multiplier Theorem

Let f and g be continuously differentiable functions of two variables. Suppose that on the level curve $g(x, y) = c$ the function f attains its extrema at (x_0, y_0). If $(\nabla g)(x_0, y_0) \neq \mathbf{0}$, then there is a real value λ such that

$$(\nabla f)(x_0, y_0) = \lambda(\nabla g)(x_0, y_0)\,.$$

Proof. First we note that (x_0, y_0) is on the level curve $g(x, y) = c$; thus $c = g(x_0, y_0)$.

Define $F(x, y) = g(x, y) - g(x_0, y_0)$. Then F has continuous first partial derivatives, and $(\nabla F)(x_0, y_0) = (\nabla g)(x_0, y_0) \neq \mathbf{0}$. Then either $F_x(x_0, y_0) \neq 0$ or $F_y(x_0, y_0) \neq 0$. Suppose that $F_y(x_0, y_0) \neq 0$. Then the Implicit Function Theorem implies that there exists $\delta > 0$ a unique differentiable function $h : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ such that

$$F(x, h(x)) = 0 \qquad \text{and} \qquad y_0 = h(x_0)\,.$$

In other words, the set $\big\{(x, h(x)) \,\big|\, x_0 - \delta < x < x_0 + \delta\big\}$ is a subset of the level curve $g(x, y) = g(x_0, y_0)$. Therefore, the function $G : (x_0 - \delta, x_0 + \delta) \to \mathbb{R}$ defined by $G(x) = f(x, h(x))$ attains its extrema at (an interior point) x_0; thus

$$G'(x_0) = f_x(x_0, y_0) + f_y(x_0, y_0)h'(x_0) = 0\,.$$

Since the implicit differentiation shows that

$$h'(x_0) = -\frac{F_x(x_0, h(x_0))}{F_y(x_0, h(x_0))} = -\frac{g_x(x_0, y_0)}{g_y(x_0, y_0)}\,,$$

we conclude that

$$f_x(x_0, y_0) - f_y(x_0, y_0)\frac{g_x(x_0, y_0)}{g_y(x_0, y_0)} = 0\,.$$

If $f_y(x_0, y_0) = 0$, then $f_x(x_0, y_0) = 0$ which implies that $(\nabla f)(x_0, y_0) = \mathbf{0} = 0 \cdot (\nabla g)(x_0, y_0)$. If $f_y(x_0, y_0) \neq 0$, then

$$\frac{f_x(x_0, y_0)}{f_y(x_0, y_0)} = \frac{g_x(x_0, y_0)}{g_y(x_0, y_0)}$$

which implies that $(\nabla f)(x_0, y_0) /\!/ (\nabla g)(x_0, y_0)$; thus there exists λ such that

$$(\nabla f)(x_0, y_0) = \lambda(\nabla g)(x_0, y_0)\,.$$

Similar argument can be applied to the case $F_x(x_0, y_0) \neq 0$, and we omit the proof for this case. □

Remark 13.77. The scalar λ in the theorem above is called a Lagrange multiplier.

Example 13.78. Find the extreme value of $f(x,y) = 4xy$ subject to the constraint

$$\frac{x^2}{9} + \frac{y^2}{16} = 1\,.$$

Let $g(x,y) = \dfrac{x^2}{9} + \dfrac{y^2}{16} - 1$. Suppose that on the level curve $g(x,y) = 0$ the function f attains its extrema at (x_0, y_0). Note that then $(\nabla g)(x_0, y_0) \neq \mathbf{0}$ (since $(x_0, y_0) \neq (0,0)$); thus the Lagrange Multiplier Theorem implies that there exists $\lambda \in \mathbb{R}$ such that

$$(4y_0, 4x_0) = (\nabla f)(x_0, y_0) = \lambda(\nabla g)(x_0, y_0) = \lambda\Big(\frac{2x_0}{9}, \frac{y_0}{8}\Big)\,.$$

Therefore, (x_0, y_0) satisfies $4y_0 = \dfrac{2\lambda x_0}{9}$ and $4x_0 = \dfrac{\lambda y_0}{8}$, as well as $\dfrac{x_0^2}{9} + \dfrac{y_0^2}{16} = 1$. Therefore, $\lambda \neq 0$, and

$$4x_0 = \frac{\lambda y_0}{8} = \frac{\lambda}{8} \cdot \frac{\lambda x_0}{18} = \frac{\lambda^2 x_0}{144}\,.$$

The identity above implies that $x_0 = 0$ or $\lambda = \pm 24$.

1. If $x_0 = 0$, then $y_0 = \pm 4$ which shows that $\lambda = 0$, a contradiction.

2. If $\lambda = \pm 24$, then $x_0 = \pm\dfrac{3y_0}{4}$; thus

$$1 = \frac{1}{9} \cdot \frac{9y_0^2}{16} + \frac{y_0^2}{16} = \frac{y_0^2}{8}\,.$$

   Therefore, $y_0 = \pm 2\sqrt{2}$ which implies that $x_0 = \pm\dfrac{3\sqrt{2}}{2}$. At these (x_0, y_0), $f(x_0, y_0) = \pm 24$. Therefore, on the ellipse $\dfrac{x^2}{9} + \dfrac{y^2}{16} = 1$ the maximum of f is 24 $\Big($at $(x_0, y_0) = \big(\pm 2\sqrt{2}, \pm\frac{3\sqrt{2}}{2}\big)\Big)$ and the minimum of f is -24 $\Big($at $(x_0, y_0) = \big(\pm 2\sqrt{2}, \mp\frac{3\sqrt{2}}{2}\big)\Big)$.

Example 13.79. Find the extreme value of $f(x,y) = 4xy$, where $x > 0$ and $y > 0$, subject to the constraint $\dfrac{x^2}{9} + \dfrac{y^2}{16} = 1$. From the previous example we find that the maximum of f is 24 $\Big($at $(x_0, y_0) = \big(2\sqrt{2}, \frac{3\sqrt{2}}{2}\big)\Big)$. The minimum of f occurs at the end-points $(0,4)$ or $(3,0)$. In either points, the value of f is 0; thus the minimum of f is 0.

Example 13.80. Find the extreme value of $f(x,y) = 4xy$, where (x,y) satisfies $\dfrac{x^2}{9}+\dfrac{y^2}{16} \leqslant 1$.
We have find the extreme value of f, under the constraint $\dfrac{x^2}{9}+\dfrac{y^2}{16} = 1$, is ± 24. Therefore, it suffices to consider the extreme value of f in the interior $\dfrac{x^2}{9}+\dfrac{y^2}{16} < 1$.

Assume that f attains its extreme value at an interior point (x_0, y_0). Then (x_0, y_0) is a critical point of f; thus

$$f_x(x_0, y_0) = f_y(x_0, y_0) = 0$$

which implies that $(x_0, y_0) = (0, 0)$. Since $f(0,0) = 0$, $f(0,0)$ is not an extreme value of f. Therefore, the extreme value of f on the region $\dfrac{x^2}{9}+\dfrac{y^2}{16} \leqslant 1$ is ± 24.

We note that $(0,0)$ in fact is a saddle point of f since $f_{xx}(0,0)f_{yy}(0,0) - f_{xy}(0,0)^2 = -16 < 0$.

Example 13.81. Find the extreme value of $f(x,y) = x^2 + 6(y^2+y+1)^2$ subject to the constraint $x^2 + (y^3-1)^2 = 1$ (using the method of Lagrange multipliers).

Let $g(x,y) = x^2 + (y^3-1)^2$. We first compute the gradient of f and g as follows:

$$(\nabla f)(x,y) = \big(2x, 12(2y+1)(y^2+y+1)\big) \quad \text{and} \quad (\nabla g)(x,y) = \big(2x, 6y^2(y^3-1)\big) .$$

Assume that f, under the constraint $g = 1$, attains its extrema at (x_0, y_0). Then

1. If $(\nabla g)(x_0, y_0) \neq \mathbf{0}$, then the Lagrange multiplier theorem implies that there exists $\lambda \in \mathbb{R}$ such that

$$\big(2x_0, 12(2y_0+1)(y_0^2+y_0+1)\big) = \lambda\big(2x_0, 6y_0^2(y_0^3-1)\big) . \tag{13.10.1}$$

   Therefore, $x_0(\lambda - 1) = 0$ and $2(2y_0+1) = \lambda y_0^2(y_0-1)$.

   (a) $x_0 = 0$, then $g(x_0, y_0) = 1$ implies that $y_0 = \sqrt[3]{2}$ ($y_0 = 0$ cannot be true because no λ will verify (13.10.1)); thus $f(x_0, y_0) = 6(\sqrt[3]{4}+\sqrt[3]{2}+1)^2$.

   (b) $\lambda = 1$, then $4y_0 + 2 = y_0^2(y_0-1)$ or equivalently, $y_0^3 - y_0^2 - 2(2y_0+1) = 0$. Note that

$$y_0^3 - y_0^2 - 4y_0 - 2 = (y_0+1)(y_0^2 - 2y_0 - 2) ;$$

   thus $y_0 = -1$ (impossible since $g(x_0, -1) \neq 1$) or $y_0 = 1 \pm \sqrt{3}$ (both are impossible since $g(x_0, 1\pm\sqrt{3}) \neq 1$).

2. If $(\nabla g)(x_0, y_0) = \mathbf{0}$, then $(x_0, y_0) = (0, 0)$; thus $f(x_0, y_0) = 1$.

Therefore, the maximum of f, under the constraint $g = 1$, is $f(0, \sqrt[3]{2}) = 6(\sqrt[3]{4} + \sqrt[3]{2} + 1)^2$ and the minimum of f, under the constraint $g = 1$, is $f(0, 0) = 1$.

Similar argument of proving Theorem 13.76 can be used to show the following

Theorem 13.82

Let f and g be continuously differentiable functions of n variables. Suppose that on the level curve $g(x_1, \cdots, x_n) = c$ the function f attains its extrema at $(a_1, \cdots, a_n)$. If $(\nabla g)(a_1, \cdots, a_n) \neq \mathbf{0}$, then there is a real value λ such that

$$(\nabla f)(a_1, \cdots, a_n) = \lambda(\nabla g)(a_1, \cdots, a_n)\,.$$

Example 13.83. Find the minimum value of $f(x, y, z) = 2x^2 + y^2 + 3z^2$ subject to the constraint $2x - 3y - 4z = 49$.

Let $g(x, y, z) = 2x - 3y - 4z - 49$. Then $(\nabla g) \neq \mathbf{0}$; thus if f attains its relative extrema at (x_0, y_0, z_0), there exists $\lambda \in \mathbb{R}$ such that $(\nabla f)(x_0, y_0, z_0) = \lambda(\nabla g)(x_0, y_0, z_0)$. Therefore,

$$(4x_0, 2y_0, 6z_0) = \lambda(2, -3, -4)$$

or equivalently, $\lambda = 2x_0 = -\dfrac{2}{3}y_0 = -\dfrac{3}{2}z_0$. Since $2x_0 - 3y_0 - 4z_0 = 49$, we find that $\lambda = 6$ which implies that

$$(x_0, y_0, z_0) = (3, -9, -4)\,.$$

Since f grows beyond any bound as $\sqrt{x^2 + y^2 + z^2}$ approaches ∞, we find that $f(3, -9, -4) = 147$ is the minimum of f.

Next, we consider the optimization problem of finding the extreme value of a function of three variables $w = f(x, y, z)$ subject to two constraints $g(x, y, z) = h(x, y, z) = 0$.

Theorem 13.84: Lagrange Multiplier Theorem - More General Version

Let f, g and h be continuously differentiable functions of three variables. Suppose that subject to the constraints $g(x, y, z) = h(x, y, z) = c$ the function f attains its extrema at (x_0, y_0, z_0). If $(\nabla g)(x_0, y_0, z_0) \times (\nabla h)(x_0, y_0, z_0) \neq \mathbf{0}$, then there are real numbers λ and μ such that

$$(\nabla f)(x_0, y_0, z_0) = \lambda(\nabla g)(x_0, y_0, z_0) + \mu(\nabla h)(x_0, y_0, z_0)\,.$$

Example 13.85. Find the extreme value of the function $f(x, y, z) = 20 + 2x + 2y + z^2$ subject to two constraints $x^2 + y^2 + z^2 = 11$ and $x + y + z = 3$.

Let $g(x, y, z) = x^2 + y^2 + z^2 - 11$ and $h(x, y, z) = x + y + z - 3$. We first note that if (x, y, z) satisfies $g(x, y, z) = h(x, y, z) = 0$, then $(\nabla g)(x, y, z) \times (\nabla h)(x, y, z) \neq \mathbf{0}$. Moreover, f attains its extrema on the intersection of the level surface $g(x, y, z) = 0$ and $h(x, y, z) = 0$ (since the intersection is closed and bounded). Suppose that f attains its extrema at (x_0, y_0, z_0). Then there exists $\lambda, \mu \in \mathbb{R}$ such that

$$\begin{aligned}(\nabla f)(x_0, y_0, z_0) &= \lambda(\nabla g)(x_0, y_0, z_0) + \mu(\nabla h)(x_0, y_0, z_0)\,,\\ g(x_0, y_0, z_0) &= h(x_0, y_0, z_0) = 0\,.\end{aligned}$$

Therefore,

$$2\lambda x_0 + \mu = 2\,, \tag{13.10.2a}$$

$$2\lambda y_0 + \mu = 2\,, \tag{13.10.2b}$$

$$2(\lambda - 1)z_0 + \mu = 0\,, \tag{13.10.2c}$$

$$x_0^2 + y_0^2 + z_0^2 = 11\,, \tag{13.10.2d}$$

$$x_0 + y_0 + z_0 = 3\,. \tag{13.10.2e}$$

(13.10.2a,b) implies that $\lambda(x_0 - y_0) = 0$; thus $\lambda = 0$ or $x_0 = y_0$.

1. If $\lambda = 0$, then (13.10.2a) implies $\mu = 2$ and (13.10.2c) implies $\mu = 2z_0$. Therefore, $z_0 = 1$ which further shows $x_0^2 + y_0^2 = 10$ and $x_0 + y_0 = 2$. Then $(x_0, y_0) = (3, -1)$ or $(-1, 3)$. Therefore, when $\lambda = 0$,

$$(x_0, y_0, z_0) = (3, -1, 1) \quad \text{or} \quad (x_0, y_0, z_0) = (-1, 3, 1)\,.$$

2. If $x_0 = y_0$, then (13.10.2d,e) implies that $2x_0^2 + z_0^2 = 11$ and $2x_0 + z_0 = 3$. Therefore, $x_0 = y_0 = \dfrac{3 \pm 2\sqrt{3}}{3}$, $z_0 = \dfrac{3 \mp 4\sqrt{3}}{3}$.

Since $f(3, -1, 1) = f(-1, 3, 1) = 25$ and

$$f\Big(\frac{3+2\sqrt{3}}{3}, \frac{3+2\sqrt{3}}{3}, \frac{3-4\sqrt{3}}{3}\Big) = f\Big(\frac{3-2\sqrt{3}}{3}, \frac{3-2\sqrt{3}}{3}, \frac{3+4\sqrt{3}}{3}\Big) = \frac{91}{3}\,,$$

we conclude that the maximum and minimum value of f subject to $g = h = 0$ are $\dfrac{91}{3}$ and 25, respectively.

Example 13.86. Find the extreme value of $f(x,y,z) = z$ subject to the constraints $x^4 + y^4 - z^3 = 0$ and $y = z$.

Let $g(x,y,z) = x^4 + y^4 - z^3$ and $h(x,y,z) = y - z$. Then

$$(\nabla g)(x,y,z) = (4x^3, 4y^3, -3z^2) \qquad \text{and} \qquad (\nabla h)(x,y,z) = (0,1,-1)$$

which implies that

$$(\nabla g)(x,y,z) \times (\nabla h)(x,y,z) = (3z^2 - 4y^3, 4x^3, 4x^3)\,.$$

Suppose the extreme value of f, under the constraints $g = h = 0$, occurs at (x_0, y_0, z_0).

1. If $(\nabla g)(x_0,y_0,z_0) \times (\nabla h)(x_0,y_0,z_0) = \mathbf{0}$, then $(x_0,y_0,z_0) = (0,0,0)$ and $f(0,0,0) = 0$.

2. If $(\nabla g)(x_0,y_0,z_0) \times (\nabla h)(x_0,y_0,z_0) \neq \mathbf{0}$, then the Lagrange Multiplier Theorem implies that there exist $\lambda, \mu \in \mathbb{R}$ such that

$$(\nabla f)(x_0,y_0,z_0) = \lambda(\nabla g)(x_0,y_0,z_0) + \mu(\nabla h)(x_0,y_0,z_0)\,.$$

Therefore, (x_0,y_0,z_0) satisfies that

$$4\lambda x_0^3 = 0\,, \tag{13.10.3a}$$
$$4\lambda y_0^3 + \mu = 0\,, \tag{13.10.3b}$$
$$-3\lambda z_0^2 - \mu = 1\,, \tag{13.10.3c}$$
$$x_0^4 + y_0^4 - z_0^3 = 0\,, \tag{13.10.3d}$$
$$y_0 - z_0 = 0\,. \tag{13.10.3e}$$

Then (13.10.3a) implies that $\lambda = 0$ or $x_0 = 0$.

(a) If $\lambda = 0$, then (13.10.3b) shows $\mu = 0$; thus using (13.10.3c), we obtain a contradiction $0 = -1$. Therefore, $\lambda \neq 0$.

(b) If $x_0 = 0$ (and $\lambda \neq 0$), then (13.10.3d) implies that $y_0^4 - z_0^3 = 0$. Together with (13.10.3e), we find that $y_0 = 0$ or $y_0 = 1$. However, if $y_0 = 0$, then (13.10.3b) shows that $\mu = 0$ which again implies a contradiction $0 = 1$ from (13.10.3c). Therefore, $y_0 = z_0 = 1$ (and there are λ, μ satisfying (13.10.3b,c) for $y_0 = z_0 = 1$ but the values of λ and μ are not important).

Therefore, the Lagrange Multiplier Theorem only provides one possible $(x_0, y_0, z_0) = (0, 1, 1)$ where f attains its extreme value.

Since the intersection of the level surface $g = 0$ and $h = 0$ is closed and bounded, f must attains its maximum and minimum subject to the constraints $g = h = 0$. Since $(0, 0, 0)$ and $(0, 1, 1)$ are the only possible points where f attains its extrema, the maximum and minimum of f, subject to the constraint $g = h = 0$, is $f(0, 1, 1) = 1$ and $f(0, 0, 0) = 0$, respectively.