Chapter 2

Differentiation

2.1 Functions of several variables

Definitions, Notations

A function has three important components.

(1) a **domain** X

(2) a **codomain** Y and

(3) **a rule of assignment** that associates to each element x in the domain X a unique element denoted by $f(x)$ in the domain Y.

Usually we use the notation $f : X \to Y$.

Definition 2.1.1. The range of a function $f : X \to Y$ is the set of all elements of Y that are actual values of f, i.e,

$$\text{range of } f = \{y \in Y | y = f(x) \text{ for some } x \in X\}.$$

Definition 2.1.2. f is called **onto** (surjective) if for every $y \in Y$, there is an element $x \in X$ such that $f(x) = y$. It is **one-to-one**(injective) if $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

In this section, we study functions whose domain is $\mathbb{R}^n$ or its subset with values in $\mathbb{R}^m$.

Definition 2.1.3. If the domain of f is $\mathbb{R}^n$ or its subset and the range is $\mathbb{R}$ or its subset, then f is called n**-variable scalar-valued function**. In particular, if $n \geq 2$, it is called **functions of several variables**. If the domain is $A \subset \mathbb{R}^n$ then we write $f\colon A \subset \mathbb{R}^n \to \mathbb{R}$. Sometimes $\mathbf{x} \mapsto f(\mathbf{x})$ is used. If the range is $\mathbb{R}^m$, $m \geq 2$ it is called **vector-valued function**. Use $f\colon \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$. We denote $f(\mathbf{x}) = (f_1, \ldots, f_m) \in \mathbb{R}^m$, where $f_i\colon \mathbb{R}^n \to \mathbb{R}$ functions of n variables. In other words, $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_n(\mathbf{x}))$ each f_i is called ***i*-th component** (***i*-th component function**) of f.

Example 2.1.4. (1) Let $L(\mathbf{x}) = \|\mathbf{x}\|$. This is the "length"function defined on $\mathbb{R}^n$

(2) Consider $N(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ for $\mathbf{x} \in \mathbb{R}^3 - \{\mathbf{0}\}$.

Visualizing functions

Definition 2.1.5. The **graph** of a functions of several variables $f\colon A \subset \mathbb{R}^n \to \mathbb{R}$ is (**graph**) the following set

$$graph(f) = \{(\mathbf{x}, f(\mathbf{x})) \in \mathbb{R}^{n+1} \mid \mathbf{x} \in A \subset \mathbb{R}^n\}.$$

Componentwise,

$$graph(f) = \{(x_1, \cdots, x_n, f(x_1, \cdots, x_n)) \in \mathbb{R}^{n+1} \mid \mathbf{x} \in A \subset \mathbb{R}^n\}.$$

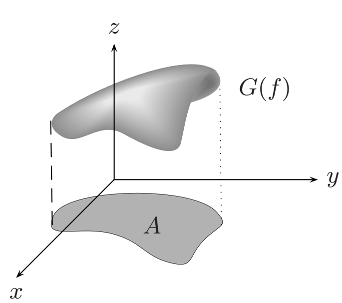


Figure 2.1: Graph of two variable function

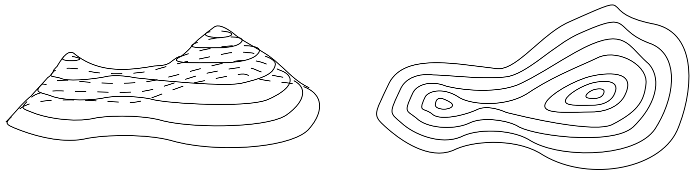


Figure 2.2: Level set

Level sets, curves, surfaces

Definition 2.1.6. The **level set** of $f\colon \mathbb{R}^n \to \mathbb{R}$ is the set of all $\mathbf{x}$ where the function f has constant value:

$$S_c = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = c, c \in \mathbb{R}\}.$$

If $n = 2$, it is **level curve** and if $n = 3$, **level surface**. x-**section** of the graph of f is the set

$$\{(x, y, z) \in \mathbb{R}^3 | z = f(x, y), x = c\}.$$

Similarly, y-**section** can be defined.

Example 2.1.7. The graph of $f(x, y) = x^2 + y^2$ is called **paraboloid** or **paraboloid of revolution**. Draw the level sets.

sol. The level set of $x^2 + y^2 = c$ is 0 if $c = 0$. For $c > 0$ it is a circle of radius $\sqrt{c}$. If $c < 0$, the level set is empty.

□

Example 2.1.8. Draw level sets of $f(x, y) = x^2 - y^2$. The graph is called **hyperbolic paraboloid** or **saddle**.

sol. The level sets of $f(x, y) = x^2 - y^2 = c$.

Detail view of the level set. If $c = 0$, then it is $y = \pm x$, two lines through origin. If $c > 0$, the level set is a hyperbola meeting with x-axis, and if $c < 0$ level set is a hyperbola meeting with y-axis. The intersection with xz-plane is the parabola $z = x^2$, and the intersection with yz-plane is the parabola $z = -y^2$. Hence the graph of f is as in Figure 2.3.

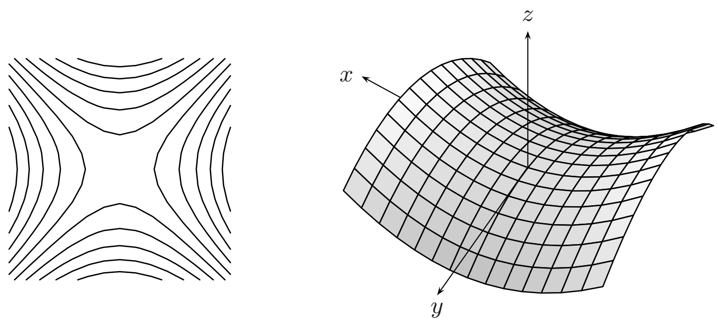
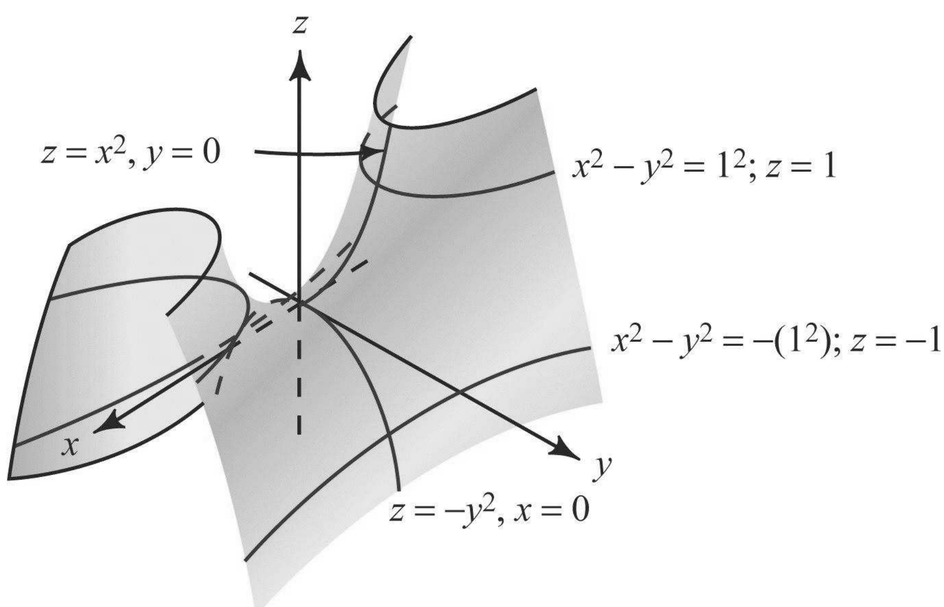


Figure 2.3: The graph of $f(x, y) = x^2 - y^2$ and its level curve

□

The graph of this function $f(x, y) = x^2 - y^2$ from the homepage of Marsden's



Level surface of function of three variables

Example 2.1.9. Study the level surface of $f(x, y, z) = x^2 + y^2 + z^2$.

$\boxed{\text{sol.}}$ The set $x^2 + y^2 + z^2 = c$ becomes

$$\begin{cases} \text{origin} & \text{if } c = 0 \\ \text{circle of radius } \sqrt{c} & \text{if } c > 0 \\ \text{empty if} & \text{if } c < 0 \end{cases}$$

To imagine the graph in $\mathbb{R}^4$, consider intersection with $\mathbb{R}^3_{z=0} = \{(x, y, z, w) \mid z = 0\}$. It is

$$\{(x, y, z, w) \mid w = x^2 + y^2, z = 0\}$$

□

Example 2.1.10. Describe the graph of $f(x, y, z) = x^2 + y^2 - z^2$.

$\boxed{\text{sol.}}$ The graph of $f = x^2 + y^2 - z^2$ is a subset of 4-dimensional space. If we denote the points in this space by (x, y, z, t), then the graph is given by

$$\{(x, y, z, t)|t = x^2 + y^2 - z^2\}.$$

The level surface is

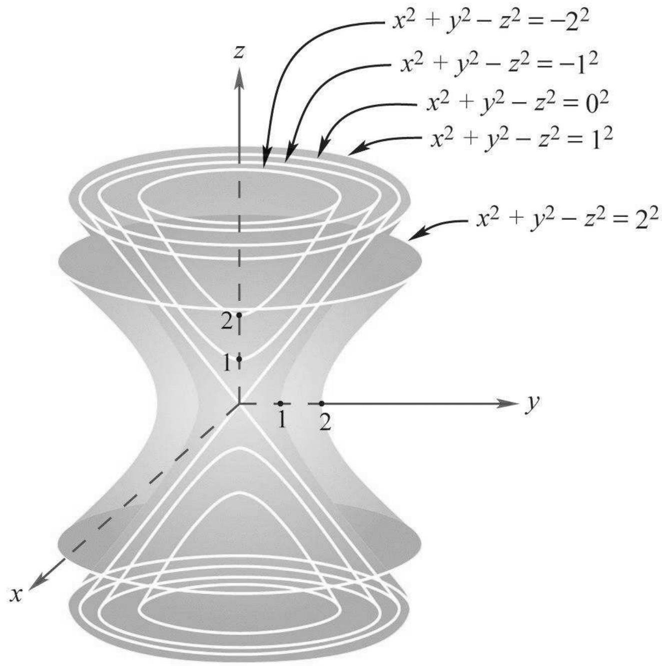
$$L_c = \{(x, y, z)|x^2 + y^2 - z^2 = c\}.$$

We have three cases:

(1) For $c = 0$, we have $z = \pm\sqrt{x^2 + y^2}$. This is a cone.

(2) If $c = -a^2$, we obtain $z = \pm\sqrt{x^2 + y^2 + a^2}$. This is a **hyperboloid of two sheets**.

(3) If $c = a^2 > 0$, we obtain $z = \pm\sqrt{x^2 + y^2 - a^2}$. This is **hyperboloid of single sheet**.



On the other hand, if we consider intersection with $y = 0$; $S_{y=0} = \{(x, y, z, t) \mid y = 0\}$, the intersection with the graph of f is

$$S_{y=0} \cap \text{ graph of } f = \{(x, y, z, t) \mid y = 0, t = x^2 - z^2\}.$$

By changing the role of y and z we have

$$\{(x, y, z, t) \mid t = x^2 - y^2, z = 0\}.$$

This set is considered to belong to (x, y, t)-space and is a hyperbolic paraboloid(saddle).

□

2.2 Limits and Continuity

Limit using ε-δ

Definition 2.2.1 (Limit suing ε-δ). Let $\mathbf{f}\colon \mathbf{A} \subset \mathbb{R}^{\mathbf{n}} \to \mathbb{R}^{\mathbf{m}}$. We say the **limit** of $\mathbf{f}$ at $\mathbf{x}_0 \in \mathbb{R}^n$ is $\mathbf{b}$, if for any $\varepsilon > 0$ there exists some positive δ such that for all $\mathbf{x} \in A$ satisfying $0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta$, the inequality $\|\mathbf{f}(\mathbf{x}) - \mathbf{b}\| < \varepsilon$ holds.

Example 2.2.2. The following function is defined at all points except $(0,0)$.

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

Find the limit as $\mathbf{x} \to (0,0)$.

$\boxed{\textbf{sol.}}$ We know in one variable calculus that

$$\lim_{x\to 0} \frac{\sin x}{x} = 1.$$

So we can guess

$$\lim_{(x,y)\to(0,0)} \frac{\sin(x^2+y^2)}{x^2+y^2} = \lim_{(x,y)\to(0,0)} \frac{\sin \|(x,y)\|^2}{\|(x,y)\|^2} = 1.$$

For any $\varepsilon > 0$, there exists $\delta > 0$ such that $|x| < \delta \implies |(\sin x)/x - 1| < \varepsilon$. Here we can assume $0 < \delta < 1$. Write $\mathbf{v} = (x,y)$. If $\|\mathbf{v}\| < \delta$ holds, then

$$|f(x,y) - 1| = \left|\frac{\sin(\|\mathbf{v}\|^2)}{\|\mathbf{v}\|^2} - 1\right| < \varepsilon.$$

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

□

Example 2.2.3. Show

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

$\boxed{\textbf{sol.}}$ Observe

$$0 \le \frac{xy}{\sqrt{x^2+y^2}} \le \frac{x^2+y^2}{\sqrt{x^2+y^2}} = \sqrt{x^2+y^2}.$$

For any ε, choose $\delta = \varepsilon$. Then for $\|(x,y) - (0,0)\| < \delta$, we have

$$\left|\frac{xy}{\sqrt{x^2+y^2}} - 0\right| = \frac{xy}{\sqrt{x^2+y^2}} \le \sqrt{x^2+y^2} = \|(x,y) - (0,0)\| < \delta = \varepsilon.$$

Thus the limit is 0.

□

Example 2.2.4. How about the following limits ?

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

and

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

sol. (a) Set $y = 0$ and let $x \to 0$. Next set $x = 0$ let $y \to 0$. The limit is different! Hence the limit does not exist.

(b) Note that

$$0 \le \frac{2x^2y}{x^2+y^2} \le \frac{2x^2y}{x^2} = 2|y|.$$

For any ε, choose $\delta = \varepsilon/2$. Then for $\|(x,y)\| < \delta$, we have

$$\left|\frac{2x^2y}{x^2+y^2} - 0\right| < 2\delta = \varepsilon.$$

Thus the limit is 0.

□

Definition 2.2.5 (Open sets)**.** Let $\mathbf{x}_0 \in \mathbb{R}^n$. The **open ball** (or disk) of radius r with center $\mathbf{x}_0$ is the set of all points $\mathbf{x}$ such that $\|\mathbf{x}-\mathbf{x}_0\| < r$. This is denoted by $B_r(\mathbf{x}_0)$ or $B(\mathbf{x}_0; r)$. A closed ball is a set of the form $\|\mathbf{x}-\mathbf{x}_0\| \le r$.

A set $U \subset \mathbb{R}^n$ is said to be **open** if for every point $\mathbf{x}_0 \in U$, there exists some $r > 0$ such that $B_r(\mathbf{x}_0)$ is contained in U(in symbol, $B_r(\mathbf{x}_0) \subset U$).

Theorem 2.2.6. $B_r(\mathbf{x}_0)$ *itself is open.*

Example 2.2.7. Half plane is open.

Definition 2.2.8 (Boundary)**.** Let $X \subset \mathbb{R}^n$. A point $\mathbf{x} \in \mathbb{R}^n$ is called a **boundary point** of X if every neighborhood of $\mathbf{x}$ contains at least a point in X and at least a point not in X. A set $X \subset \mathbb{R}^n$ is said to be **closed** if it contains all of its boundary points. Finally, a **neighborhood** of a point $x \in X$ is an open set containing x and contained in X.

Let us define the concept of a limit using open sets.

Definition 2.2.9 (Limit)**.** Suppose $f\colon X \subset \mathbb{R}^n \to \mathbb{R}^m$, where $X \subset \mathbb{R}^n$ and let $\mathbf{a} \in X$ or boundary of X. Then the meaning of

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \mathbf{L}$$

is as follows: Given any $\epsilon > 0$, we can find a $\delta > 0$ such that if points $\mathbf{x} \in X$ are in the open ball of radius δ centered at $\mathbf{a}$, then the point $f(\mathbf{x})$ remain inside of an open ball of radius ϵ centered at $\mathbf{L}$.

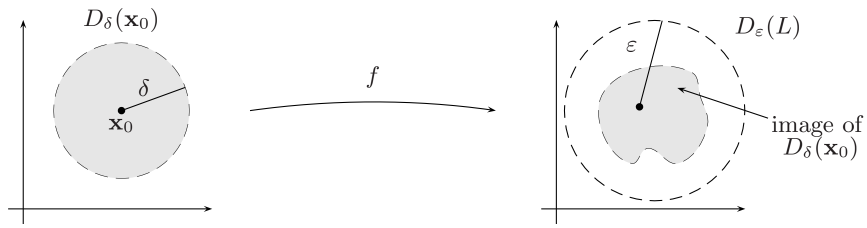


Figure 2.4: Limit using neighborhood

Example 2.2.10. Let $f : \mathbb{R}^2 - \mathbf{0}$ be defined by

$$f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}. \;\; or \; (\frac{x^2 - y^4}{x^2 + y^4})$$

Study the behavior near the origin.

$\boxed{\textbf{sol.}}$ This function is undefined at $\mathbf{0} = (0, 0)$. First observe

$$f(x, 0) = \frac{x^2}{x^2} = 1.$$

But

$$f(0, y) = \frac{-y^2}{y^2} = -1.$$

Hence limit cannot exists. In general, we can check along the line $y = mx$

$$f(x, y) = \frac{x^2 - m^2x^2}{x^2 + m^2x^2} = \frac{1 - m^2}{1 + m^2}.$$

Thus the limit does not exist.

□

This example might lead the student to believe the limit exists if the limit along straight lines exists and equal. But this is not true. See the following example (Exer 23)

$$f(x,y) = \frac{x^4y^4}{(x^2+y^4)^3}.$$

Along any line $y = ax$ the limit is zero. While along $y = a\sqrt{x}$, we see the limit is $\frac{a^4}{(1+a^4)^3}$.

Properties of Limits

Theorem 2.2.11. *Let $\mathbf{F}, \mathbf{G} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ be vector valued functions and $f, g : X \subset \mathbb{R}^n \to \mathbb{R}$ be scalar-valued functions and let $\mathbf{a}$ be a point of X or boundary and k a scalar. Assume $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{L}$, $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{G}(\mathbf{x}) = \mathbf{M}$ and $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = L$, $\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = M$. Then the following hold:*

(1) $\lim_{\mathbf{x}\to\mathbf{a}} (\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) = \mathbf{L} + \mathbf{M}$.

(2) $\lim_{\mathbf{x}\to\mathbf{a}} k\mathbf{F}(\mathbf{x}) = k\mathbf{L}$

(3) $\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x})g(\mathbf{x})) = LM$.

(4) $\lim_{\mathbf{x}\to\mathbf{a}} \dfrac{f(\mathbf{x})}{g(\mathbf{x})} = \dfrac{L}{M}$, $(M \neq 0)$.

Using this theorem, we can find limits of polynomials or rational functions.

Example 2.2.12. Find the limit of the following functions.

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

(2) $\lim_{(x,y)\to(0,0)} \dfrac{x^2 + xy^3 - x^2y + 2}{xy + 3}$

sol. By above theorem,

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

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

□

Theorem 2.2.13. *Let $\mathbf{f} = (f_1, \cdots, f_m) : X \subset \mathbb{R}^n \to \mathbb{R}^m$ be a vector valued function. Then $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ if and only if $\lim_{\mathbf{x}\to\mathbf{a}} f_i(\mathbf{x}) = L_i$, for $i = 1, \cdots, m$.*

Continuity

Definition 2.2.14. $\mathbf{f}\colon X \subset \mathbb{R}^n \to \mathbb{R}^m$ **continuous** at $\mathbf{a} \in X$ if

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

We say $\mathbf{f}$ is continuous on X if it is so at all points of X.

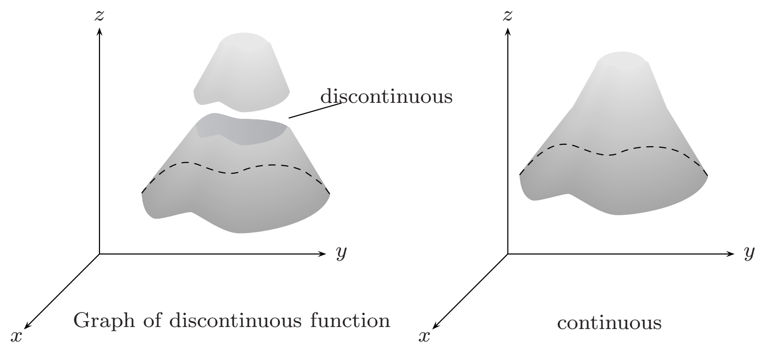


Figure 2.5: continuous, discontinuous function

Example 2.2.15. Show that the following function is continuous at $(0,0)$.

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

$\boxed{\textbf{sol.}}$ We have seen in example 2.2.3 that the limit of this function at the origin is 0, and it equals the functon value $f(0,0)$. Hence f is continuous there.

□

Theorem 2.2.16 (Composite function)**.** *Suppose* $\mathbf{g}\colon A \subset \mathbb{R}^n \to \mathbb{R}^m$ *and* $\mathbf{f}\colon B \subset \mathbb{R}^m \to \mathbb{R}^p$ *are given Suppose* $g(A) \subset B$ *so that* $\mathbf{f} \circ g$ *is defined. If* $\mathbf{g}$ *is continuous at* $\mathbf{x}_0 \in A$ *and* $\mathbf{f}(\mathbf{x}_0) \in B$*, and* $\mathbf{f}$ *is continuous at* $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0)$*, then* $\mathbf{f} \circ \mathbf{g}$ *is continuous at* $\mathbf{x}_0$*.*

Example 2.2.17. Show $f(x,y) = \cos^2((y+x^3)/(1+x^2))$ is continuous.

2.3 The Derivative

Partial derivatives

Recall 1-dim. case 'Differentiable' means 'smooth' in some sense: At least tangent line must be defined. Also the composite of differentiable functions is differentiable.

Definition 2.3.1. Let $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ be a real valued function. Then the **partial derivative** with respect to i-th variable x_i is:

$$\lim_{h\to 0} \frac{f(x_1,\dots,x_i+h,\dots,x_n) - f(x_1,\dots,x_n)}{h} = \lim_{h\to 0} \frac{f(\mathbf{x}+h\mathbf{e}_i) - f(\mathbf{x})}{h}$$

whenever the limit exists. The partial derivative of f with respect to x_i at $\mathbf{x}_0 \in \mathbb{R}^n$ is denoted by

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_0), \quad \text{or} \quad \left.\frac{\partial f}{\partial x_i}\right|_{\mathbf{x}_0}.$$

For vector valued function $\mathbf{f}\colon U \subset \mathbb{R}^n \to \mathbb{R}^m$, the partial derivative is the partial derivative of each component function f_j, where $\mathbf{f} = (f_1,\dots,f_m)$.

Example 2.3.2. Find partial derivatives of $f(x,y) = x^2y + \cos(x+y)$.

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

□

Example 2.3.3. Find partial derivatives of $g(x,y) = xy/\sqrt{x^2+y^2}$ at $(1,1)$.

sol. First we compute $\frac{\partial g}{\partial x}(1,1)$:

$$\begin{aligned}\frac{\partial g}{\partial x}(1,1) &= \frac{y\sqrt{x^2+y^2} - xy(x/\sqrt{x^2+y^2})}{x^2+y^2} \\ &= \frac{y(x^2+y^2) - x^2y}{(x^2+y^2)^{3/2}} \\ &= 2^{3/2}.\end{aligned}$$

□

Example 2.3.4. Find partial derivatives at $(0,0)$ of the function defined by

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

sol. Use definition:

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

□

Linear approximation

Suppose f is a scalar valued function of two variables. We want to find the equation of tangent plane to the graph of $z = f(x,y)$ at (a,b). Let us fix $y = b$ and consider the y-section:

$$z = f(x,b).$$

The direction vector of the tangent line to this curve is $(1, 0, f_x(a,b))$. To compute the parametric equation of the tangent line to this curve in point-direction form, we choose the point $(a, b, f(a,b))$ and the direction vector $\mathbf{u} := (1, 0, f_x(a,b))$. Thus the equation is

$$\ell_1(t) = (a, b, f(a,b)) + t(1, 0, f_x(a,b)).$$

Similarly, we consider the x- section $z = f(a,y)$. For this curve we obtain a direction vector $\mathbf{v} := (0, 1, f_x(a,b))$ and

$$\ell_2(t) = (a, b, f(a,b)) + t(0, 1, f_y(a,b)).$$

Now the plane containing these two tangent lines are determined by the point $(a, b, f(a,b))$ and the normal vector $\mathbf{N} = \mathbf{u}\times\mathbf{v}$. By computation, we see

$$\mathbf{u}\times\mathbf{v} = -f_x(a,b)\mathbf{i} - f_y(a,b)\mathbf{j} + \mathbf{k}.$$

The slope along x-direction is $\partial f/\partial x(a,b)$ and the slope along y-direction is $\partial f/\partial y(a,b)$. Since the point $(a, b, f(a,b))$ lies in the plane, we see $(x-a, y-b, z-f(a,b)) \perp \mathbf{N}$. Hence

$$\boxed{z = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b).}$$

This can be interpreted as follows: The equation of the plane can be written as $z = A(x-a) + B(y-b) + f(a,b)$. Here the slope along y-section is A which must be $\frac{\partial f}{\partial x}(a,b)$, while the slope along x-section is $\frac{\partial f}{\partial y}(a,b)$ which is B.

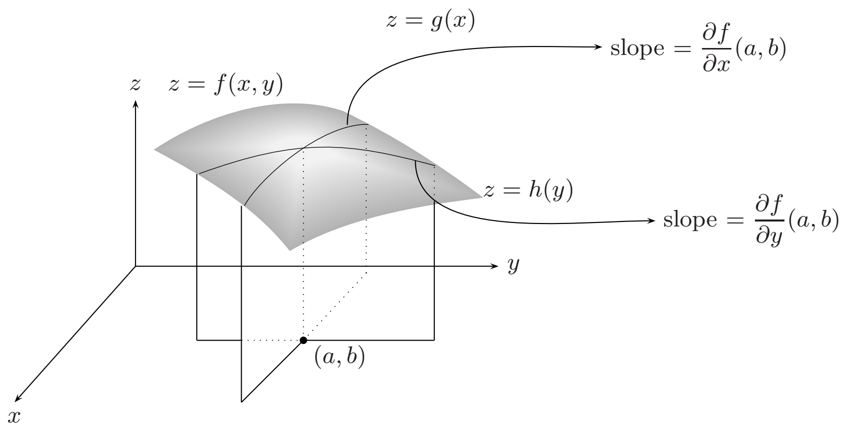


Figure 2.6: Geometric meaning of partial derivative

Example 2.3.5. Find partial derivative of $f(x,y) = x^{1/3}y^{1/3}$ by definition,

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

Similarly $(\partial f/\partial y)(0,0) = 0$. But this is not differentiable, as we shall see later.

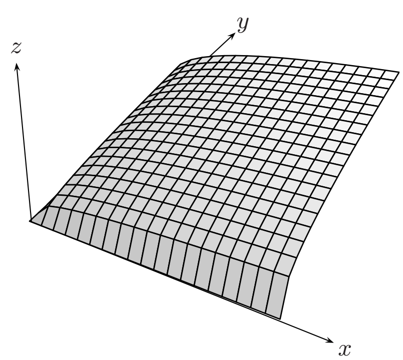


Figure 2.7: Graph of $f(x,y) = x^{1/3}y^{1/3}$

Existence of partial derivatives is not enough to say a function has a tangent plane

Example 2.3.6. Let $y = x$ in the above example. Then $f(x,x) = x^{2/3}$ is not differentiable at 0 in one variable sense. Hence there is no *tangent line* at $(0,0)$ in one variable sense. Then the *tangent plane* of course does not exist!

Differentiation of a function of several variable

Review: A one variable differentiable function $f(x)$ can be approximated near a point a by tangent line at x: $f(a) + f'(a)(x-a)$. It is the **tangent line approximation** or **linear approximation** of $f(x)$. It satisfies (figure 2.8)

$$\lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x-a)}{x-a} = 0. \tag{2.1}$$

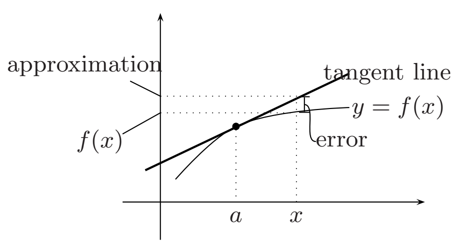


Figure 2.8: tangent approximation of a function of one variable

Now consider a two variable function $f\colon \mathbb{R}^2 \to \mathbb{R}$. Assume f_x and f_y exist at (a,b). Then one can consider the following plane:

$$z = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b). \tag{2.2}$$

Ask a question: Is this a tangent plane to the surface $z = f(x,y)$? The answer is NOT NECESSARILY! A tangent plane is defined only when the plane given by (2.2) approximates $f(x,y)$ in a sense similar to (2.1). i.e., we have the following definition.

Definition 2.3.7. We say $f\colon \mathbb{R}^2 \to \mathbb{R}$ **differentiable** at (a,b) if $\partial f/\partial x$ and

$\partial f/\partial y$ exists and for $(x,y)\to(a,b)$, the limit

$$\frac{f(x,y)-f(a,b)-\dfrac{\partial f}{\partial x}(a,b)(x-a)-\dfrac{\partial f}{\partial y}(a,b)(y-b)}{\|(x,y)-(a,b)\|}\to 0.$$

In this case the **tangent plane** at (a,b) is given by

$$z=f(a,b)+\frac{\partial f}{\partial x}(a,b)(x-a)+\frac{\partial f}{\partial y}(a,b)(y-b).$$

If a function is differentiable at all points of its domain, we say it is **differentiable**.

Definition 2.3.8. In general, Suppose $f\colon\mathbb{R}^n\to\mathbb{R}$. Then we say f differentiable at $\mathbf{a}$ if

$$\lim_{\mathbf{x}\to\mathbf{a}}\frac{f(\mathbf{x})-f(\mathbf{a})-\left[\frac{\partial f}{\partial x_1}(\mathbf{a}),\cdots,\frac{\partial f}{\partial x_n}(\mathbf{a})\right]\begin{bmatrix}x_1-a_1\\ \cdots\\ x_n-a_n\end{bmatrix}}{\|\mathbf{x}-\mathbf{a}\|}=0.$$

In short,

$$\boxed{\lim_{\mathbf{x}\to\mathbf{a}}\frac{f(\mathbf{x})-f(\mathbf{a})-\mathbf{D}f(\mathbf{a})(\mathbf{x}-\mathbf{a})}{\|\mathbf{x}-\mathbf{a}\|}=0.}\tag{2.3}$$

Here $\mathbf{D}f(\mathbf{a})=[\frac{\partial f}{\partial x_1}(\mathbf{a}),\cdots,\frac{\partial f}{\partial x_n}(\mathbf{a})]$ is called the **derivative of** f. Usually the derivative of a scalar function is written as a row vector(for the convenience of matrix operations.)

Example 2.3.9. Show $f(x,y)=x^2+y^2$ is differentiable at $(0,0)$. Find the tangent plane of $f(x,y)=x^2+y^2$ at $(0,0)$.

sol. We see $(\partial f/\partial x)(0,0)=(\partial f/\partial y)(0,0)=0$.

$$\begin{aligned}&\lim_{(x,y)\to(0,0)}\frac{f(x,y)-f(0,0)-\dfrac{\partial f}{\partial x}(0,0)(x)-\dfrac{\partial f}{\partial y}(0,0)(y)}{\|(x,y)-(0,0)\|}\\ &=\lim_{(x,y)\to(0,0)}\frac{f(x,y)}{\|(x,y)\|}=\lim_{(x,y)\to(0,0)}\sqrt{x^2+y^2}\\ &=0.\end{aligned}$$

Hence it is differentiable at $(0,0)$. The tangent plane is $z=0$.

□

Example 2.3.10. Show the function defined by

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

is differentiable at $(0,0)$.

$\boxed{\textbf{sol.}}$ It is easy to see that $f_x(0,0) = f_y(0,0) = 0$ by definition. Now

$$\begin{aligned}
&\lim_{(x,y)\to(0,0)} \frac{f(x,y) - f(0,0) - 0 - 0}{\|(x,y) - (0,0)\|} = \lim_{(x,y)\to(0,0)} \frac{f(x,y)}{\|(x,y)\|} \\
&= \lim_{(x,y)\to(0,0)} \frac{2x^2y^2}{(x^2+y^2)^{3/2}} \le \lim_{(x,y)\to(0,0)} \frac{xy(x^2+y^2)}{(x^2+y^2)^{3/2}} \\
&= \lim_{(x,y)\to(0,0)} \frac{xy}{(x^2+y^2)^{1/2}} \\
&\le \lim_{(x,y)\to(0,0)} \frac{x^2+y^2}{2(x^2+y^2)^{1/2}} = 0.
\end{aligned}$$

□

Example 2.3.11. Find a tangent plane to $z_1 = x^2 - xy + y^2$ which is parallel to any of tangent plane to the surface $z_2 = x^2 + y$.

$\boxed{\textbf{sol.}}$ The normal vector to the surface $z_1 = x^2 - xy + y^2$ is $(2x - y, -x + 2y, -1)$, while the normal vector to $z_2 = x^2 + y$ is $(2x, 1, -1)$. Assume they are parallel. Then

$$(2x - y, -x + 2y, -1) = k(2x, 1, -1)$$

From this we obtain $k = 1, x = -1, y = 0, z_1 = 1, z_2 = 1$. Thus the equation of tangent plane to the first surface is $z - 1 = -2(x + 1) - y$.

□

Differentiability of vector valued function

Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a vector valued function with several variables. If every component of $\mathbf{f}$ is differentiable, we say $\mathbf{f}$ is differentiable. We can express the concept of differentiability of a vector function in vector notation as follows:

Definition 2.3.12. A function $\mathbf{f} = (f_1, \dots, f_m)\colon \mathbb{R}^n \to \mathbb{R}^m$ is said to be **differentiable** at a point $\mathbf{a}$ if all the partial derivatives of $\mathbf{f}$ exists at $\mathbf{a}$ and

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - \mathbf{Df}(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} = 0$$

holds. Here

$$\mathbf{Df}(\mathbf{a}) = \begin{bmatrix} \mathbf{D}f_1 \\ \vdots \\ \mathbf{D}f_m \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}$$

is a $m \times n$ matrix and $\mathbf{Df}(\mathbf{a})(\mathbf{x}-\mathbf{a})$ is the product of $m \times n$ matrix $\mathbf{D}f(\mathbf{a})$ and the $n \times 1$ vector $\mathbf{x} - \mathbf{a}$. $\mathbf{Df}(\mathbf{a})$ is called the **derivative** of f at $\mathbf{a}$. Sometimes it is called the **Jacobian matrix**. Also, note that this is a vector version of (2.3).

If $m = 1$, then

$$\mathbf{D}f = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} & \cdots & \dfrac{\partial f}{\partial x_n} \end{bmatrix}.$$

It is also called the **gradient** of f and denoted by ∇f. If we let $\mathbf{h} = \mathbf{x} - \mathbf{a}$, then real valued function f is differentiable at a point $\mathbf{a}$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{1}{\|\mathbf{h}\|} \left| f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\mathbf{a}) h_j \right| = 0.$$

Example 2.3.13. Find the derivative of $\mathbf{Df}(x, y)$.

(1) $\mathbf{f}(x, y) = (xy, x + y)$

(2) $\mathbf{f}(x, y) = (e^{x+y}, x^2 + y^2, xe^y)$

sol. (1) $f_1 = xy,\ f_2 = x + y$. Hence

$$\mathbf{Df}(\mathbf{x}) = \begin{bmatrix} y & x \\ 1 & 1 \end{bmatrix}.$$

(2) $f_1 = e^{x+y},\ f_2 = x^2 + y^2,\ f_3 = xe^y$. Hence

$$\mathbf{Df}(\mathbf{x}) = \begin{bmatrix} e^{x+y} & e^{x+y} \\ 2x & 2y \\ e^y & xe^y \end{bmatrix}.$$

□

Example 2.3.14. Show $f(x,y) = (xy, x+y)$ is differentiable at $(0,0)$.

sol. From example 2.3.13,

$$\mathbf{Df}(0,0) = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

$$\begin{aligned} &\lim_{(x,y)\to(0,0)} \frac{\left\| \mathbf{f}(x,y) - \mathbf{f}(0,0) - \mathbf{Df}(0,0) \begin{bmatrix} x \\ y \end{bmatrix} \right\|}{\|(x,y)-(0,0)\|} \\ &= \lim_{(x,y)\to(0,0)} \frac{\|(xy, x+y) - (0, x+y)\|}{\|(x,y)\|} \\ &= \lim_{(x,y)\to(0,0)} \frac{|xy|}{\sqrt{x^2+y^2}} = 0. \end{aligned}$$

□

Relation with continuity

Theorem 2.3.15. *If $\mathbf{f} = (f_1, \ldots, f_n) \colon \mathbb{R}^n \to \mathbb{R}^m$ has all partial derivatives $\partial f_i / \partial x_j$ exist and continuous in a neighborhood of $\mathbf{x}$, then $\mathbf{f}$ is differentiable at $\mathbf{x}$.*

Example 2.3.16. $\mathbf{f}(x,y) = (e^{xy}, x^2+y^2, xe^y)$ is differentiable at all points of $\mathbb{R}^2$.

sol. Since all the partial derivatives are continuous on $\mathbb{R}^2$, f is differentiable by Theorem 2.3.15.

□

Example 2.3.17. Given

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

Show that

(1) The partial derivatives at $(0,0)$ exist.

(2) f is not differentiable at $(0,0)$.

sol. (1) From definition, we have

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

and

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

(2) We have

$$\mathbf{D}f = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}\Big|_{(0,0)} = \begin{bmatrix} 0 & 0 \end{bmatrix}.$$

Hence we consider the following limit:

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

Since $\lim_{(x,y)\to(0,0)} xy/(x^2+y^2)$ does not exists, f is not differentiable at $(0,0)$.

□

But if we change the f to $\dfrac{x^2y}{\sqrt{x^2+y^2}}$ in the above example, then we can show it is differentiable at $(0,0)$.

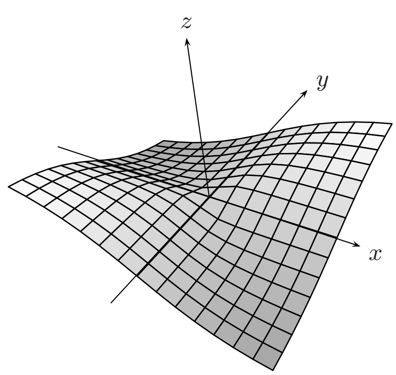


Figure 2.9: Graph of example 2.3.17

Theorem 2.3.18. *If* $\mathbf{f} = (f_1, \ldots, f_n)\colon \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\mathbf{x}_0$*, then* $\mathbf{f}$ *is continuous at* $\mathbf{x}_0$*.*

The converse is not true.

Example 2.3.19. Suppose $f\colon \mathbb{R}^2 \to \mathbb{R}$ is given as follows.

$$f(x,y) = \begin{cases} 1 & x = 0 \text{ or } y = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{tikzcd}
\text{partial deriv. exist} \arrow[dd, Rightarrow, shift left, "/" marking] & & \\
 & \text{differentiable} \arrow[ul, Rightarrow] \arrow[dl, Rightarrow] & \text{continuous partial derivative} \arrow[l, Rightarrow] \\
\text{continuous} \arrow[uu, Rightarrow, shift left, "/" marking] & &
\end{tikzcd}$$

2.4 Higher order derivatives; Newton's Method

Properties of derivatives

Proposition 2.4.1 (Rules)**.** *Suppose* $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ *and* $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\mathbf{x}_0$*. Then we have*

(1) ***[constant multiple rule]*** *For all constant* c, $c\mathbf{f}$ *is differentiable at* $\mathbf{x}_0$.

$$\mathbf{D}(c\mathbf{f})(\mathbf{x}_0) = c\mathbf{D}\mathbf{f}(\mathbf{x}_0)$$

(2) ***[sum rule]*** *Sum* $\mathbf{f} + \mathbf{g}$ *differentiable at* $\mathbf{x}_0$

$$\mathbf{D}(\mathbf{f} + \mathbf{g})(\mathbf{x}_0) = \mathbf{D}\mathbf{f}(\mathbf{x}_0) + \mathbf{D}\mathbf{g}(\mathbf{x}_0)$$

(3) ***[product rule]*** *(When* $m = 1$*) Product* fg *differentiable at* $\mathbf{x}_0$.

$$\mathbf{D}(fg)(\mathbf{x}_0) = g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0) + f(\mathbf{x}_0)\mathbf{D}g(\mathbf{x}_0)$$

(4) ***[quotient rule]*** *(When* $m = 1$*) If* $g(\mathbf{x}_0) \neq 0$, *then* f/g *differentiable at* $\mathbf{x}_0$.

$$\mathbf{D}\left(\frac{f}{g}\right)(\mathbf{x}_0) = \frac{g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0) - f(\mathbf{x}_0)\mathbf{D}g(\mathbf{x}_0)}{(g(\mathbf{x}_0))^2}$$

Rule (1) and (2) together is called the "linearity".

Proof. (3) Suppose $\mathbf{x} \to \mathbf{x}_0$. We need to show that

$$\frac{g(\mathbf{x})f(\mathbf{x}) - g(\mathbf{x}_0)f(\mathbf{x}_0) - [g(\mathbf{x})\mathbf{D}f(\mathbf{x}_0) + f(\mathbf{x})\mathbf{D}g(\mathbf{x}_0)](\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|} \to 0.$$

First we see the numerator:

$$\begin{aligned}
g(\mathbf{x})f(\mathbf{x}) - g(\mathbf{x})f(\mathbf{x}_0) + g(\mathbf{x})f(\mathbf{x}_0) - g(\mathbf{x}_0)f(\mathbf{x}_0)& \\
- [g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0) + f(\mathbf{x}_0)\mathbf{D}g(\mathbf{x}_0)](\mathbf{x} - \mathbf{x}_0)& \\
= [g(\mathbf{x})f(\mathbf{x}) - g(\mathbf{x})f(\mathbf{x}_0) - g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)]& \\
+ [g(\mathbf{x})f(\mathbf{x}_0) - g(\mathbf{x}_0)f(\mathbf{x}_0) - f(\mathbf{x}_0)\mathbf{D}g(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)]&.
\end{aligned}$$

Let A be the terms in the first bracket and B be the terms in the second bracket. Then

$$\begin{aligned}
A &= g(\mathbf{x})f(\mathbf{x}) - g(\mathbf{x})f(\mathbf{x}_0) - g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \\
&= g(\mathbf{x})f(\mathbf{x}) - g(\mathbf{x})f(\mathbf{x}_0) - g(\mathbf{x})\mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \\
&\qquad + g(\mathbf{x})\mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) - g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \\
&= g(\mathbf{x})[f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)] + [g(\mathbf{x}) - g(\mathbf{x}_0)]\mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0).
\end{aligned}$$

Similar expression for B. Now using the definition of derivative and continuity we see

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{A}{\|\mathbf{x} - \mathbf{x}_0\|} = 0, \qquad \lim_{\mathbf{x} \to \mathbf{x}_0} \frac{B}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

□

Example 2.4.2. $n = 2$, $m = 3$. Let $\mathbf{f} = (xe^y, x \sin xy, e^{x+y})$ and $\mathbf{g} = (x + y, x^2 + e^y, x^2 - y)$. Then

$$D\mathbf{f} = \begin{bmatrix} e^y & xe^y \\ \sin xy + xy \cos xy & x^2 \cos xy \\ e^{x+y} & e^{x+y} \end{bmatrix}, \quad D\mathbf{g} = \begin{bmatrix} 1 & 1 \\ 2x & e^y \\ 2x & -1 \end{bmatrix}.$$

Example 2.4.3 (Product rule)**.** Let $f = e^{xy}$ and $g = x + y^2$. Then $fg = e^{xy}(x + y^2)$ and

$$D(fg) = \left[y(x + y^2)e^{xy} + e^{xy}, x(x + y^2)e^{xy} + 2ye^{xy} \right]$$

while

$$gDf + fDg = (x + y^2)[ye^{xy}, xe^{xy}] + e^{xy}[1, 2y].$$

Higher order partial derivatives

Suppose f has $\partial f/\partial x$, $\partial f/\partial y$ and each of these partials again has partial derivatives. Then we write

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x}$$
$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2}.$$

These are the **second partial derivatives** of f. We also use simplified expressions such as $\partial f/\partial x = f_x$, $\partial f/\partial y = f_y$, and

$$\frac{\partial^2 f}{\partial x^2} = f_{xx}, \quad \frac{\partial^2 f}{\partial y \partial x} = f_{xy}, \quad \frac{\partial^2 f}{\partial x \partial y} = f_{yx}, \quad \frac{\partial^2 f}{\partial y^2} = f_{yy}.$$

Example 2.4.4. Compute higher order partial derivatives.

(1) $f(x, y) = e^{xy} + x^2 y$

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

$\boxed{\textbf{sol.}}$ (1) For $f(x, y) = e^{xy} + x^2 y$,

$$f_x = ye^{xy} + 2xy, \qquad f_y = xe^{xy} + x^2$$
$$f_{xx} = y^2 e^{xy} + 2y, \qquad f_{yy} = x^2 e^{xy}$$
$$f_{xy} = f_{yx} = e^{xy} + xye^{xy} + 2x.$$

(2) For $f(x, y, z) = x^2 y + y^2 z + z^3 x$,

$$f_x = 2xy + z^3, \quad f_y = x^2 + 2yz, \quad f_z = y^2 + 3z^2 x$$
$$f_{xx} = 2y, \qquad f_{yy} = 2z, \qquad f_{zz} = 6xz$$
$$f_{xy} = f_{yx} = 2x$$
$$f_{xz} = f_{zx} = 3z^2$$
$$f_{yz} = f_{zy} = 2y.$$

□

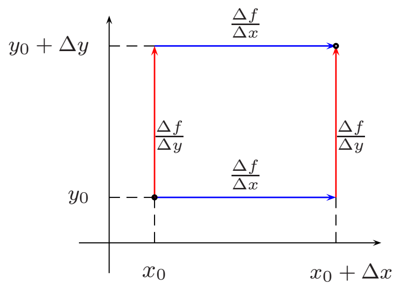


Figure 2.10: Two paths of difference quotient

Mixed partial derivatives

Definition 2.4.5. Assume $U \subset \mathbb{R}^n$ is open. A scalar valued function $f : U \to \mathbb{R}$ is said to be class $\mathcal{C}^k$ if all partial derivatives up to order k exist and are continuous.

Derivatives such as $f_{xy}, f_{yx},\ f_{xz}, f_{zx},\ f_{yz}, f_{zy}$ are called **mixed partial derivatives**.

Theorem 2.4.6. *If $f(x, y)$ is class $\mathcal{C}^2$, then*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

This theorem holds for functions with several variables, i.e., if $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is class $\mathcal{C}^2$, then

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i} \text{ for } i, j = 1, \cdots, n.$$

Proof. A motivation. Recall for fixed y, $\frac{\partial f}{\partial x}(x, y)$ is the limit of

$$\frac{\Delta_x f}{\Delta x} = \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}. \tag{2.4}$$

Then the mixed derivative $\frac{\partial^2 f}{\partial y \partial x}$ must be the limit of $\frac{\Delta_y(\frac{\partial f}{\partial x})}{\Delta y}$. Hence let us compute the increment along y in (2.4) and take the limit

$$\frac{\Delta_y(\frac{\Delta_x f}{\Delta x})}{\Delta y} = \frac{f(x+\Delta x, y+\Delta y) - f(x, y+\Delta y) - [f(x+\Delta x, y) - f(x,y)]}{\Delta y \Delta x} \tag{2.5}$$

as $\Delta x, \Delta y \to 0$(in some sense). Here Δ_x is the increment of f along x defined by

$$\Delta_x f(x,y) = f(x+\Delta x, y) - f(x,y).$$

and Δ_y is the increment of f along y defined by

$$\Delta_y f(x,y) = f(x, y+\Delta y) - f(x,y).$$

Now we show that indeed the limit of (2.5) $(\Delta x, \Delta y) \to 0$ is $\frac{\partial^2 f}{\partial y \partial x}$.

$$\begin{aligned}
\Delta_y[\Delta_x f(x,y)] &= \Delta_y[f(x+\Delta x, y) - f(x,y)] \\
&= \Delta_y f(x+\Delta x, y) - \Delta_y f(x,y) \\
&= f(x+\Delta x, y+\Delta y) - f(x+\Delta x, y) - [f(x, y+\Delta y) - f(x,y)].
\end{aligned}$$

Note this difference is taken along the blue line first, then along red line in the figure. By changing the order, we see

$$\begin{aligned}
\Delta_x[\Delta_y f(x,y)] &= \Delta_x[f(x, y+\Delta y) - f(x,y)] \\
&= \Delta_x f(x, y+\Delta y) - \Delta_x f(x,y) \\
&= f(x+\Delta x, y+\Delta y) - f(x, y+\Delta y) - [f(x+\Delta x, y+\Delta y) - f(x,y)] \\
&= \Delta_y[\Delta_x f(x,y)].
\end{aligned}$$

Let's define the increment of f(correspond. to second derivative) by

$$\begin{aligned}
S(\Delta x, \Delta y) = f(x+\Delta x, y+\Delta y) &- f(x+\Delta x, y) \\
&- f(x, y+\Delta y) + f(x,y).
\end{aligned}$$

Fix y and Δy and define

$$g(x) = f(x, y+\Delta y) - f(x,y).$$

Then $S(\Delta x, \Delta y) = g(x+\Delta x) - g(x)$. Use MVT. There is $x \le \tilde{x} \le x + \Delta x$

such that $S(\Delta x, \Delta y) = g'(\tilde{x})\Delta x$. So

$$\begin{aligned} S(\Delta x, \Delta y) &= g'(\tilde{x})\Delta x \\ &= \Big(\frac{\partial f}{\partial x}(\tilde{x}, y + \Delta y) - \frac{\partial f}{\partial x}(\tilde{x}, y)\Big)\Delta x. \end{aligned}$$

Apply MVT for y again there is $y \le \tilde{y} \le y + \Delta y$ such that

$$S(\Delta x, \Delta y) = \frac{\partial^2 f}{\partial y \partial x}(\tilde{x}, \tilde{y})\Delta x \Delta y.$$

Since $\partial^2 f/\partial x \partial y$ is continuous, we have

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = \lim_{(\Delta x, \Delta y)\to(0,0)} \frac{S(\Delta x, \Delta y)}{\Delta x \Delta y}.$$

Here the l.h.s. depends on the order of x, y but r.h.s. is independent of x and y. Now exchanging the role of x and y we see

$$\frac{\partial^2 f}{\partial x \partial y}(x, y) = \lim_{(\Delta x, \Delta y)\to(0,0)} \frac{S(\Delta y, \Delta x)}{\Delta y \Delta x} = \lim_{(\Delta x, \Delta y)\to(0,0)} \frac{S(\Delta x, \Delta y)}{\Delta x \Delta y} = \frac{\partial^2 f}{\partial y \partial x}(x, y).$$

□

Example 2.4.7. Find mixed partial $\partial^2 f/\partial y \partial x$ of

$$f(x, y) = xy^2 - e^{x^2 - x}/(x^2 + 1).$$

sol. By Thm 2.4.6, f_{yx} may be computed instead of f_{xy}.

$$f_{yx} = (2xy)_x = 2y.$$

Note that for this particular example, computing f_{yx} is simpler than f_{xy} .

□

Example 2.4.8. Show the following when f belongs to $\mathcal{C}^3$.

$$\frac{\partial^3 f}{\partial x \partial y \partial z} = \frac{\partial^3 f}{\partial z \partial x \partial y}.$$

sol. We can change order of differentiation

$$\begin{aligned}\frac{\partial^3 f}{\partial x\partial y\partial z} &= \frac{\partial}{\partial x}\Big(\frac{\partial^2 f}{\partial y\partial z}\Big) = \frac{\partial}{\partial x}\Big(\frac{\partial^2 f}{\partial z\partial y}\Big)\\ &= \frac{\partial^2}{\partial x\partial z}\Big(\frac{\partial f}{\partial y}\Big) = \frac{\partial^2}{\partial z\partial x}\Big(\frac{\partial f}{\partial y}\Big)\\ &= \frac{\partial^3 f}{\partial z\partial x\partial y}.\end{aligned}$$

□

2.5 Chain rule

Chain rule in several variables

Theorem 2.5.1 (Chain rule-simple)**.** *Suppose* $\mathbf{x}(t) = (x(t), y(t))\colon \mathbb{R} \to \mathbb{R}^2$ *differentiable at* t_0 *and* $f\colon X \subset \mathbb{R}^2 \to \mathbb{R}$ *differentiable at* $\mathbf{x}_0 = \mathbf{x}(t_0)$ *then the composite* $h = f \circ \mathbf{x}\colon \mathbb{R} \to \mathbb{R}$ *(* $h(t) = f(x(t), y(t))$*) is differentiable at* t_0 *and its derivative* $dh/dt(t_0)$ *is*

$$\frac{dh}{dt}(t_0) = \frac{\partial f}{\partial x}(\mathbf{x}_0)\frac{dx}{dt}(t_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)\frac{dy}{dt}(t_0).$$

Proof. From

$$\frac{dh}{dt}(t_0) = \lim_{t\to t_0}\frac{h(t)-h(t_0)}{t-t_0},$$

we have

$$\begin{aligned}\frac{h(t)-h(t_0)}{t-t_0} &= \frac{f(x(t),y(t)) - f(x(t_0),y(t_0))}{t-t_0}\\ &= \frac{f(x(t),y(t)) - f(x(t_0),y(t)) + f(x(t_0),y(t)) - f(x(t_0),y(t_0))}{t-t_0}.\end{aligned}$$

Since f differentiable in each variable, we see by mean value theorem that there exists a point c between $x(t)$ and $x(t_0)$ such that

$$f(x(t),y(t)) - f(x(t_0),y(t)) = \Big(\frac{\partial f}{\partial x}(c, y(t))\Big)(x(t) - x(t_0))$$

holds. Similarly,

$$\frac{h(t)-h(t_0)}{t-t_0} = \Big(\frac{\partial f}{\partial x}(c, y(t))\Big)\frac{x(t)-x(t_0)}{t-t_0} + \Big(\frac{\partial f}{\partial y}(x(t_0), d)\Big)\frac{y(t)-y(t_0)}{t-t_0}.$$

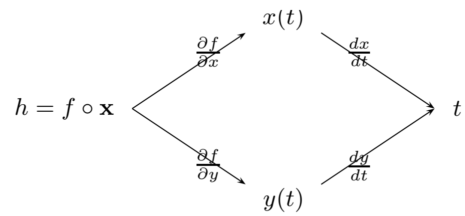


Figure 2.11: Chain rule-simple case

Let t approach t_0. Then we obtain the result. □

This theorem can be generalized to the case when f has several variables. Suppose $f\colon X \subset \mathbb{R}^n \to \mathbb{R}$ differentiable and $\mathbf{x}(t) = (x_1(t), \cdots, x_n(t))\colon \mathbb{R} \to \mathbb{R}^n$, then the composite function $h(t) = f \circ \mathbf{x}\colon \mathbb{R} \to \mathbb{R}$ has derivative

$$\frac{dh}{dt}(t_0) = \frac{\partial f}{\partial x_1}(\mathbf{x}_0)\frac{dx_1}{dt}(t_0) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}_0)\frac{dx_n}{dt}(t_0) = Df(\mathbf{x}_0) \cdot \mathbf{x}'(t_0).$$

Example 2.5.2. Show Chain rule holds for $f(x, y) = e^{xy}$ and $\mathbf{x}(t) = (t^2, 2t)$.

sol. Since $h(t) = f \circ \mathbf{x}(t) = f(x(t), y(t)) = e^{2t^3}$, we have $dh/dt = 6t^2 e^{2t^3}$. On the other hand, by chain rule, we have

$$\frac{dh}{dt} = ye^{xy} \cdot 2t + xe^{xy} \cdot 2 = 6t^2 e^{2t^3}.$$

□

Theorem 2.5.3. *Suppose $\mathbf{x}\colon X \subset \mathbb{R}^2 \to \mathbb{R}^3$ is given by $\mathbf{x}(t_1, t_2) = (x(t_1, t_2), y(t_1, t_2), z(t_1, t_2))$ and $f\colon \mathbb{R}^3 \to \mathbb{R}$ are differentiable mappings such that the range of $\mathbf{x}$ is contained in the domain of f. Then the composite function $h = f \circ \mathbf{x}$ is differentiable. For example, the derivative at $\mathbf{t}_0 = (t_1^0, t_2^0)$ is $\mathbf{D}h(\mathbf{t}_0) = \begin{bmatrix} \partial h/\partial t_1(\mathbf{t}_0), & \partial h/\partial t_2(\mathbf{t}_0) \end{bmatrix}$ where*

$$\begin{aligned} \frac{\partial h}{\partial t_1}(\mathbf{t}_0) &= \frac{\partial f}{\partial x}(\mathbf{x}_0)\frac{\partial x}{\partial t_1}(\mathbf{t}_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)\frac{\partial y}{\partial t_1}(\mathbf{t}_0) + \frac{\partial f}{\partial z}(\mathbf{x}_0)\frac{\partial z}{\partial t_1}(\mathbf{t}_0) \\ \frac{\partial h}{\partial t_2}(\mathbf{t}_0) &= \frac{\partial f}{\partial x}(\mathbf{x}_0)\frac{\partial x}{\partial t_2}(\mathbf{t}_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)\frac{\partial y}{\partial t_2}(\mathbf{t}_0) + \frac{\partial f}{\partial z}(\mathbf{x}_0)\frac{\partial z}{\partial t_2}(\mathbf{t}_0). \end{aligned}$$

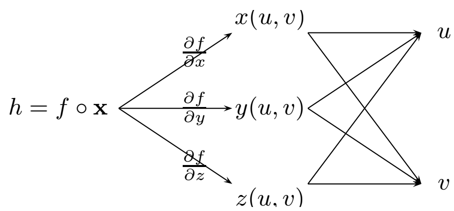


Figure 2.12: Chain rule- in Theorem 2.5.3

In matrix form, we have $\mathbf{D}h = \mathbf{D}f \circ \mathbf{D}\mathbf{x}$, where

$$\mathbf{D}f = [\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}], \quad \text{and } \mathbf{D}\mathbf{x} = \begin{bmatrix} \frac{\partial x}{\partial t_1}, \frac{\partial x}{\partial t_2} \\ \frac{\partial y}{\partial t_1}, \frac{\partial y}{\partial t_2} \\ \frac{\partial z}{\partial t_1}, \frac{\partial z}{\partial t_2} \end{bmatrix}.$$

Proof. Fix t_2 as a constant and compute $\partial h/\partial t_1$ using one variable chain rule:

$$\frac{\partial h}{\partial t_1}(\mathbf{t}_0) = \frac{\partial f}{\partial x}(\mathbf{x}_0)\frac{\partial x}{\partial t_1}(\mathbf{t}_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)\frac{\partial y}{\partial t_1}(\mathbf{t}_0) + \frac{\partial f}{\partial z}(\mathbf{x}_0)\frac{\partial z}{\partial t_1}(\mathbf{t}_0).$$

Similarly,

$$\frac{\partial h}{\partial t_2}(\mathbf{t}_0) = \frac{\partial f}{\partial x}(\mathbf{x}_0)\frac{\partial x}{\partial t_2}(\mathbf{t}_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)\frac{\partial y}{\partial t_2}(\mathbf{t}_0) + \frac{\partial f}{\partial z}(\mathbf{x}_0)\frac{\partial z}{\partial t_2}(\mathbf{t}_0).$$

□

More Generally we have: Suppose $\mathbf{x}\colon \mathbb{R}^n \to \mathbb{R}^m$ differentiable at $\mathbf{t}_0$ and $f\colon \mathbb{R}^m \to \mathbb{R}$ differentiable at $\mathbf{x}_0 = \mathbf{x}(\mathbf{t}_0)$. Then the composite function $h = f \circ \mathbf{x}$ differentiable at $\mathbf{t}_0$ and

$$\mathbf{D}h(\mathbf{t}_0) = \mathbf{D}f(\mathbf{x}_0)\mathbf{D}\mathbf{x}(\mathbf{t}_0).$$

Or

$$\begin{aligned}\frac{\partial h}{\partial t_j} &= \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \cdots + \frac{\partial f}{\partial x_m}\frac{\partial x_m}{\partial t_j} \\ &= \sum_{k=1}^{m} \frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial t_j}, \quad j = 1, \cdots n.\end{aligned}$$

Example 2.5.4. Verify the chain rule for $f(u,v,w) = u^2 + v^2 - w$, where

$$u(x,y,z) = x^2 y, \quad v = y^2, \quad z = e^{-xz}.$$

$\boxed{\textbf{sol.}}$ Let

$$h(x,y,z) = f(u(x,y,z), v(x,y,z), w(x,y,z)).$$

Then by chain rule

$$\begin{aligned} \frac{\partial h}{\partial x} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial f}{\partial w}\frac{\partial w}{\partial x} \\ &= 2u(2xy) + 2v \cdot 0 + (-1)(-ze^{-xz}). \end{aligned}$$

Other terms such as $\frac{\partial h}{\partial y}$, $\frac{\partial h}{\partial z}$ can be computed similarly.

□

Theorem 2.5.5 (Chain rule-General case)**.** *Suppose* $\mathbf{x} \colon \mathbb{R}^n \to \mathbb{R}^m$ *differentiable at* $\mathbf{t}_0$ *and* $\mathbf{f} \colon \mathbb{R}^m \to \mathbb{R}^p$ *differentiable at* $\mathbf{x}_0 = \mathbf{x}(\mathbf{t}_0)$*. Then* $\mathbf{h} = \mathbf{f} \circ \mathbf{x}$ *differentiable at* $\mathbf{t}_0$ *and*

$$\mathbf{Dh}(\mathbf{t}_0) = \mathbf{Df}(\mathbf{x}_0)\mathbf{Dx}(\mathbf{t}_0).$$

Proof. Suppose $\mathbf{x} \colon \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{f} \colon \mathbb{R}^m \to \mathbb{R}^p$ are given. Then

$$\begin{aligned} \mathbf{x}(t_1, \cdots, t_n) &= (x_1(t_1, \cdots, t_n), x_2(t_1, \cdots, t_n), \cdots, x_m(t_1, \cdots, t_n)) \\ \mathbf{f}(x_1, \cdots, x_m) &= (f_1(x_1, \cdots, x_m), \cdots, f_p(x_1, \cdots, x_m)). \end{aligned}$$

Let the composite function be

$$\mathbf{h}(t_1, \cdots, t_n) = (\mathbf{f} \circ \mathbf{x})(t_1, \cdots, t_n).$$

Apply the simple case to each component of $\mathbf{h} = [f_1 \circ \mathbf{x}, \cdots, f_p \circ \mathbf{x}]^T$(Column vector) so that $h_i = f_i \circ \mathbf{x}$ and

$$\begin{aligned} Dh_1 &= Df_1 \circ D\mathbf{x} \\ Dh_2 &= Df_2 \circ D\mathbf{x} \\ &= \cdots \\ Dh_p &= Df_p \circ D\mathbf{x}. \end{aligned}$$

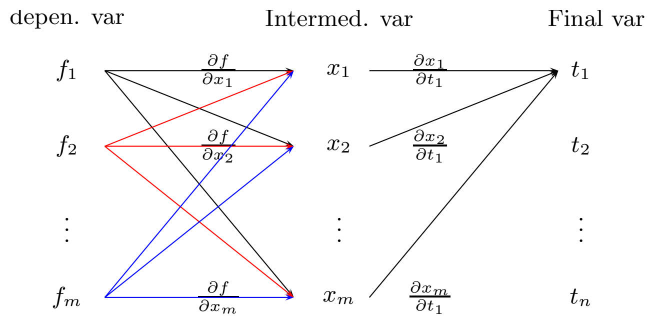


Figure 2.13: Chain rule- General Case

Now just write it in a matrix form to see $\mathbf{Dh}(\mathbf{t}_0) = \mathbf{Df}(\mathbf{x}_0)\mathbf{Dx}(\mathbf{t}_0)$. □

Example 2.5.6. Given the vector functions $\mathbf{f}$, $\mathbf{g}$ below, consider the composite function $\mathbf{h} = (k, l) = \mathbf{f} \circ \mathbf{g}$. Find the partials $\partial k/\partial x$ and $\partial l/\partial y$.

$$\mathbf{g}(x, y, z) = (xyz, x^2 + y^2 + z^2, e^{xyz}), \quad \mathbf{f}(u, v, w) = (u^2 - uv, u + v + w).$$

$\boxed{\textbf{sol.}}$ Use chain rule

$$\begin{aligned}
\frac{\partial k}{\partial x} &= \frac{\partial k}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial k}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial k}{\partial w}\frac{\partial w}{\partial x} \\
&= (2u - v)(yz) + (-u)(2x) + 0 \\
&= (2xyz - x^2 - y^2 - z^2)(yz) - (xyz)(2x) \\
&= 2xyz - 3x^2yz - y^3 - yz^2, \\
\frac{\partial l}{\partial y} &= \frac{\partial l}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial l}{\partial v}\frac{\partial v}{\partial y} + \frac{\partial l}{\partial w}\frac{\partial w}{\partial y} \\
&= 1 \cdot \frac{\partial u}{\partial y} + 1 \cdot \frac{\partial u}{\partial y} + 1 \cdot \frac{\partial w}{\partial y} \\
&= xz + 2y + xze^{xyz}.
\end{aligned}$$

Check it using matrix product.

□

Example 2.5.7. Use Chain rule to find the derivative of composite function

$$\mathbf{h}(t) = (h_1(t), h_2(t), h_2(t)) = \mathbf{f} \circ \mathbf{g}(t),$$

where $\mathbf{g}(t) = (x(t), y(t), z(t))$ and $\mathbf{f}: = (f_1(x,y,z), f_2(x,y,z), f_3(x,y,z))$.

Note that $h_i(t) = f_i(\mathbf{g}(t))$. Use Chain rule for special case(to each component)

$$\frac{dh_i}{dt} = \frac{\partial f_i}{\partial x}\frac{dx}{dt} + \frac{\partial f_i}{\partial y}\frac{dy}{dt} + \frac{\partial f_i}{\partial z}\frac{dz}{dt}.$$

Use Chain rule as a whole

$$D\mathbf{f} = \begin{bmatrix} \frac{\partial f_1}{\partial x}, & \frac{\partial f_1}{\partial y}, & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x}, & \frac{\partial f_2}{\partial y}, & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x}, & \frac{\partial f_3}{\partial y}, & \frac{\partial f_3}{\partial z} \end{bmatrix} \text{ while } D\mathbf{g} = \begin{bmatrix} x'(t) \\ y'(t) \\ z'(t) \end{bmatrix}.$$

$$\text{Hence } D\mathbf{f} \circ D\mathbf{g} = \begin{bmatrix} \frac{\partial f_1}{\partial x}x'(t) + \frac{\partial f_1}{\partial y}y'(t) + \frac{\partial f_1}{\partial z}z'(t) \\ \frac{\partial f_2}{\partial x}x'(t) + \frac{\partial f_2}{\partial y}y'(t) + \frac{\partial f_2}{\partial z}z'(t) \\ \frac{\partial f_3}{\partial x}x'(t) + \frac{\partial f_3}{\partial y}y'(t) + \frac{\partial f_3}{\partial z}z'(t) \end{bmatrix}.$$

Example 2.5.8. Let $\mathbf{f} : U \subset \mathbb{R}^n \to \mathbb{R}^m$ be given by $\mathbf{f} = (f_1, \cdots, f_m)$ and $g(\mathbf{x}) = \sin[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})]$. Compute $Dg(\mathbf{x})$.

sol.

$$Dg(\mathbf{x}) = \cos[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})]D[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})].$$

We compute $D[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})]$ which is

$$\begin{aligned} Dh &= \left[2f_1\tfrac{\partial f_1}{\partial x_1} + \cdots + 2f_m\tfrac{\partial f_m}{\partial x_1}, \cdots, 2f_1\tfrac{\partial f_1}{\partial x_n} + \cdots + 2f_m\tfrac{\partial f_m}{\partial x_n}\right] \\ &= 2\mathbf{f}(\mathbf{x})D\mathbf{f}(\mathbf{x}), \end{aligned}$$

where $D\mathbf{f}(\mathbf{x})$ is the derivative of $\mathbf{f}$, Finally, we see $Dg(\mathbf{x}) = 2\cos[\mathbf{f}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})]\mathbf{f}(\mathbf{x})D\mathbf{f}(\mathbf{x})$.

□

Example 2.5.9 (Polar/Rectangular coordinates conversions)**.** Recall

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta. \end{cases}$$

Suppose $w = f(x,y)$ is given. We would like view it as a function of (r,θ), i.e,

$$w = g(r,\theta) := f(x(r,\theta), y(r,\theta))$$

and compute $\frac{\partial g}{\partial r}, \frac{\partial g}{\partial \theta}$. By the chain rule,

$$Dg(r,\theta) = Df(x,y)D\mathbf{x}(r,\theta).$$

Hence

$$\begin{aligned} \begin{bmatrix} \frac{\partial g}{\partial r} & \frac{\partial g}{\partial \theta} \end{bmatrix} &= \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix} \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}. \end{aligned}$$

Entry wise, we see

$$\begin{cases} \frac{\partial w}{\partial r} &= \cos\theta \frac{\partial w}{\partial x} + \sin\theta \frac{\partial w}{\partial y} \\ \frac{\partial w}{\partial \theta} &= -r\sin\theta \frac{\partial w}{\partial x} + r\cos\theta \frac{\partial w}{\partial y}. \end{cases} \tag{2.6}$$

If we extract the derivative symbol only, we get a differential operator:

$$\begin{cases} \frac{\partial}{\partial r} &= \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y} \\ \frac{\partial}{\partial \theta} &= -r\sin\theta \frac{\partial}{\partial x} + r\cos\theta \frac{\partial}{\partial y}. \end{cases} \tag{2.7}$$

Similarly, we can show

$$\begin{cases} \frac{\partial}{\partial x} &= \cos\theta \frac{\partial}{\partial r} - \frac{\sin\theta}{r} \frac{\partial}{\partial \theta} \\ \frac{\partial}{\partial y} &= \sin\theta \frac{\partial}{\partial r} + \frac{\cos\theta}{r} \frac{\partial}{\partial \theta}. \end{cases} \tag{2.8}$$

2.6 Gradient and the directional derivatives

Gradient

Definition 2.6.1. Let $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable. The **gradient** of at $\mathbf{x}_0$ is

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n} \right).$$

The directional derivative

Definition 2.6.2. Let $\mathbf{v} \in \mathbf{R}^n$ be a unit vector and $\mathbf{a} \in X \subset \mathbf{R}^n$, the **directional derivative** of $f\colon X \to \mathbb{R}$ at $\mathbf{a}$ along $\mathbf{v}$ is $\mathbf{D_v}f(\mathbf{a})$ defined by

(Fig 2.14)

$$\frac{d}{dt} f(\mathbf{a}+t\mathbf{v})\Big|_{t=0}.$$

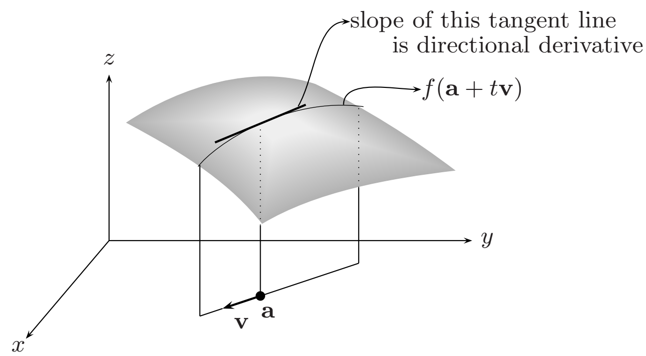


Figure 2.14: Directional Derivative

Theorem 2.6.3. *If $f(\mathbf{x})\colon X \subset \mathbb{R}^3 \to \mathbb{R}$ is differentiable and $\mathbf{a} \in X$, then the directional derivative of f at $\mathbf{a}$ along $\mathbf{v}$ exists and is given by*

$$D_{\mathbf{v}} f(\mathbf{a}) = \operatorname{grad} f(\mathbf{a}) \cdot \mathbf{v} = \nabla f(\mathbf{a}) \cdot \mathbf{v}.$$

Proof. Let $\mathbf{c}(t) = \mathbf{a} + t\mathbf{v}$ so that $f(\mathbf{a}+t\mathbf{v}) = f(\mathbf{c}(t))$. Then by the chain rule $\frac{d}{dt} f(\mathbf{c}(t)) = \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t)$. Hence

$$\frac{d}{dt} f(\mathbf{a}+t\mathbf{v})\Big|_0 = \nabla f(\mathbf{a}) \cdot \mathbf{v} = \mathbf{D} f \cdot \mathbf{c}'(t). \tag{2.9}$$

□

Example 2.6.4. The converse of above theorem does not hold. The existence of directional derivatives does not guarantee differentiability.

Consider

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

This function is not continuous at the origin, but has directional derivatives along any direction.(Fill out detail)

The directional derivative is the rate of change of f along $\mathbf{v}$. The rate of change of f along a curve is given as

$$\frac{d}{dt} f(\mathbf{c}(t))\Big|_0 = \nabla f \cdot \mathbf{c}'(t)\big|_0 = \nabla f(\mathbf{a}) \cdot \mathbf{v}. \tag{2.10}$$

Example 2.6.5. Compute the rate of change of $f(x, y, z) = xy - z^2$ at $(1, 0, 1)$ along $(1, 1, 1)$.

$\boxed{\text{sol.}}$ The unit vector to $(1, 1, 1)$ is $\mathbf{v} = (1/\sqrt{3})(1, 1, 1)$. The gradient of f at $(1, 0, 1)$ is

$$\begin{aligned}\nabla f(1, 0, 1) &= (f_z, f_y, f_z)|_{(1,0,1)} = (y, x, 2z)|_{(1,0,1)} \\ &= (0, 1, -2) \cdot \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = -\frac{1}{\sqrt{3}}.\end{aligned}$$

□

Direction of steepest ascent

We have just seen $\nabla f(\mathbf{a}) \cdot \mathbf{v}$ is the rate of change of f at $\mathbf{a}$ along the direction $\mathbf{v}$. We see

$$\mathbf{D_v} f(\mathbf{a}) = \mathbf{D} f(\mathbf{a}) \cdot \mathbf{v} = \|\mathbf{v}\| \|\nabla f(\mathbf{a})\| \cos\theta.$$

Here θ is the angle between $\mathbf{v}$ and $\nabla f(\mathbf{a})$. Hence if $\theta = 0$, the directional derivative has maximum value $\|\nabla f(\mathbf{a})\|$, and if $\theta = \pi$ it has minimum value $-\|\nabla f(\mathbf{a})\|$. Also, if $\theta = \pi/2$ then the directional derivative is 0. Hence we have

Theorem 2.6.6. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$. Then f increases fastest at $\mathbf{a}$ along $\nabla f(\mathbf{a})$. Also, f does not change along the perpendicular direction to $\nabla f(\mathbf{a})$.*

Similarly, f decreases fastest in the direction of $-\nabla f(\mathbf{a})$.

Example 2.6.7. In what direction from $(0, 1)$ does $f(x, y) = x^2 - y^2$ increases fastest?

Gradient is normal to the level set

We refer to the figure 2.15 or 2.16. Consider the level set(surface) of $f(x, y, z)$:

$$S = \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = k\}.$$

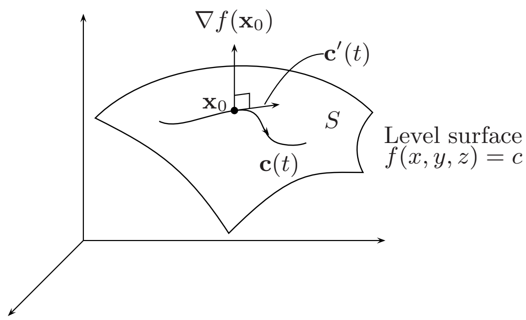


Figure 2.15: gradient at $\mathbf{x}_0$ is perpendicular to tangent plane through $\mathbf{x}_0$

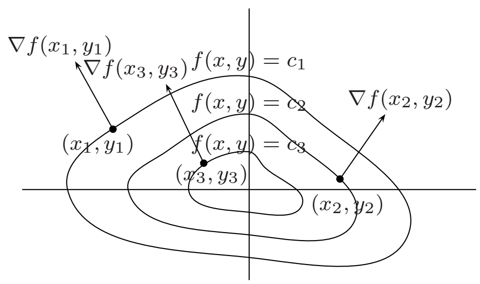


Figure 2.16: 2D case; gradient is perpendicular to level curve.

Suppose a curve $\mathbf{c}$ passes the point $\mathbf{x}_0 = (x_0, y_0, z_0)$ lies on the surface S. Then $f(\mathbf{c}(t)) = k$ holds. Then we have by chain rule

$$0 = \frac{d}{dt} f(\mathbf{c}(t)) = \nabla f(\mathbf{c}(t)) \cdot \mathbf{c}'(t).$$

Hence the tangent vector $\mathbf{c}'(t_0)$ at $\mathbf{x}_0$ is normal to the gradient $\nabla f(\mathbf{x}_0)$.

Theorem 2.6.8. *Suppose $f(x, y, z)$ is differentiable and $\nabla f(\mathbf{x}_0) \neq 0$. Then $\nabla f(\mathbf{x}_0)$ is normal to the level surface $S = \{(x, y, z) \in \mathbb{R}^3 \mid f(x, y, z) = k\}$.*

Definition 2.6.9. The **tangent plane** plane to the surface S in Theorem 2.6.8 at $\mathbf{x}_0 = (x_0, y_0, z_0)$ is given by

$$\nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) = 0, \qquad \text{or}$$
$$\frac{\partial f}{\partial x}(\mathbf{x}_0)(x - x_0) + \frac{\partial f}{\partial y}(\mathbf{x}_0)(y - y_0) + \frac{\partial f}{\partial z}(\mathbf{x}_0)(z - z_0) = 0.$$

Compare this with the definition earlier for the graph of $z = f(x, y)$:

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

Example 2.6.10. Find the equation of the tangent plane to $3xy + z^2 = 4$ at $(1, 1, 1)$.

$\boxed{\textbf{sol.}}$ The gradient $-\nabla f = (3y, 3x, 2z)$ at $(1, 1, 1)$ is $(3, 3, 2)$. Thus the tangent plane is

$$(3, 3, 2) \cdot (x - 1, y - 1, z - 1) = 0.$$

□

Example 2.6.11. Find the equation of the tangent hyper-plane to the hypersurface $x^2 + y^2 + z^2 - w^2 = 1$ at $(1, 1, -1, \sqrt{2})$.

$\boxed{\textbf{sol.}}$ The gradient $(2x, 2y, 2z, -2w)$ at $(1, 1, -1, \sqrt{2})$ it is $(2, 2, -2, 2\sqrt{2})$. Thus the tangent plane is

$$2(x - 1) + 2(y - 1) - 2(z + 1) - 2\sqrt{2}(w - \sqrt{2}) = 0.$$

□

Example 2.6.12. Consider a tangent plane to the surface

$$xyz^2 + e^{y^2 z} - \sin(x + z) - 3 = 0.$$

This surface cannot be described as the graph of a function. However, given any particular point, we may locally solve for z in terms of x, y. Thus this relation defines a function implicitly at least near such point. The question is when?

As we shall see, it is possible when $\frac{\partial F}{\partial z} \neq 0$ at some point.

Theorem 2.6.13 (Implicit function theorem)**.** *Let $F : X \subset \mathbb{R}^n \to \mathbb{R}$ be class $\mathcal{C}^1$ and let $\mathbf{a}$ be a point of the level set $S = \{\mathbf{x} \in \mathbb{R}^n | F(\mathbf{x}) = c\}$. If $F_{x_n}(\mathbf{a}) \neq 0$, then there is a neighborhood U of $(a_1, a_2, \cdots, a_{n-1})$ in $\mathbb{R}^{n-1}$ and a neighborhood V of a_n in $\mathbb{R}$, and a function $f : U \subset \mathbb{R}^{n-1} \to V$ of class $\mathcal{C}^1$ such that $x_n = f(x_1, x_2, \cdots, x_{n-1})$.*

Example 2.6.14. Consider ellipsoid $x^2/4 + y^2/36 + z^2/9 = 1$. It is the level set of the function

$$F(x, y, z) = \frac{x^2}{4} + \frac{y^2}{36} + \frac{z^2}{9}.$$

At $(\sqrt{2}, \sqrt{6}, \sqrt{3})$, we can check $\frac{\partial F}{\partial z} \neq 0$. Hence z can be solved as function of x and y.

Example 2.6.15. Let $F(x, y, z) = x^2z^2 - y$ and S be the level set of height 0. For points where $F_x = 2z^2 \neq 0$ one can solve for other variables.

Implicit function theorem-one variable

Theorem 2.6.16. *Let $y = f(x)$ and $f'(x_0) \neq 0$. Then the inverse function $x = f^{-1}(y)$ exists near $y_0 = f(x_0)$.*

We generalize this to higher dimensions. A special case first.

Theorem 2.6.17. *Suppose $F : \mathbb{R}^{n+1} \to \mathbb{R}$ has all continuous partials. Denote any point in $\mathbb{R}^{n+1}$ by $(\mathbf{x}, z)$. Assume $F(\mathbf{x}, z)$ satisfies*

$$F(\mathbf{x}_0, z_0) = 0 \text{ and } \frac{\partial F}{\partial z}(\mathbf{x}_0, z) \neq 0.$$

Then there is a ball U containing $\mathbf{x}_0$ such that $z = g(\mathbf{x})$ for $\mathbf{x} \in U$ and satisfies

$$F(\mathbf{x}, g(\mathbf{x})) = 0. \tag{2.11}$$

Moreover,

$$Dg(\mathbf{x}) = -\frac{1}{\frac{\partial F}{\partial z}|_{(\mathbf{x},g(\mathbf{x}))}} D_{\mathbf{x}}F(\mathbf{x}, z)|_{(\mathbf{x},g(\mathbf{x}))}.$$

Proof. Sketch only. Assume z is a function of $\mathbf{x}$ near $\mathbf{x}_0$.(i.e., assume there exists a function $z = g(\mathbf{x})$ satisfying (2.11). This is a big step!). Differentiating $F(\mathbf{x}, z) = 0$ with resp. to x_i, we obtain

$$\frac{\partial F}{\partial x_i} + \frac{\partial F}{\partial z}\frac{\partial z}{\partial x_i} = 0.$$

Since $F_z \neq 0$, we have for $i = 1, \cdots, n$

$$\frac{\partial z}{\partial x_i} = -\frac{\frac{\partial F}{\partial x_i}}{\frac{\partial F}{\partial z}}.$$

Hence Dg is well -defined. □

Example 2.6.18. Show that near $(x, y, u, v) = (1, 1, 1, 1)$ we can solve the following system for (u, v).

$$\begin{aligned} xu + yvu^2 &= 2 \\ xu^3 + y^2v^4 &= 2. \end{aligned}$$

$\boxed{\textbf{sol.}}$ Let

$$\begin{aligned} F_1 &= xu + yvu^2 - 2 \\ F_2 &= xu^3 + y^2v^4 - 2 \end{aligned}$$

and check that

$$\Delta = \begin{vmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{vmatrix} = \begin{vmatrix} x + 2yuv & yu^2 \\ 3u^2 & 4y^2v^3 \end{vmatrix} = \begin{vmatrix} 3 & 1 \\ 3 & 4 \end{vmatrix} = 9.$$

Hence by the implicit function theorem, we can solve it.

□

We introduce a convention: for a generic point in $\mathbf{z} \in \mathbb{R}^{n+m}$, we write $\mathbf{z} = (\mathbf{x}, \mathbf{y})$, where $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$.

Theorem 2.6.19 (Implicit function theorem, General Case)**.** *Let* $\mathbf{F} : A \subset \mathbb{R}^{n+m} \to \mathbb{R}^m$ *be class* $\mathcal{C}^1$ *where* A *is an open set in* $\mathbb{R}^{n+m}$*. Let* $(\mathbf{a}, \mathbf{b}) = (a_1, \cdots, a_n, b_1, \cdots, b_m) \in A$ *satisfy* $\mathbf{F}(\mathbf{a}, \mathbf{b}) = \mathbf{c}$*. If*

$$det\, D_{\mathbf{y}}\mathbf{F} = \left.\frac{\partial(f_1, \cdots, f_m)}{\partial(y_1, \cdots, y_m)}\right|_{(\mathbf{a},\mathbf{b})} \neq 0,$$

then there is a neighborhood U *of* $\mathbf{a}$ *in* $\mathbb{R}^n$ *and a function* $\mathbf{f} : U \to \mathbb{R}^m$ *of class* $\mathcal{C}^1$ *such that* $\mathbf{F}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{c}$ *and the derivative of* $\mathbf{f}$ *is given by*

$$D_{\mathbf{x}}\mathbf{f} = -D_{\mathbf{y}}\mathbf{F}^{-1}D_{\mathbf{x}}\mathbf{F}.$$

In other words, $\mathbf{b}$ *can be solved as a differentiable function of* $\mathbf{a}$ *in a neighborhood of* $\mathbf{a}$.