Math 2010 Chapter 7

7.1 Differentiability, Gradient

Theorem 7.1. *If $f, g : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ are differentiable at $\vec{a} \in \Omega$. Then:*

1. $f(\vec{x}) \pm g(\vec{x}), cf(\vec{x}), f(\vec{x})g(\vec{x})$ *are differentiable at* $\vec{a}$.

2. $\dfrac{f(\vec{x})}{g(\vec{x})}$ *is differentiable at* $\vec{a}$ *if* $g(\vec{a}) \neq 0$.

3. *(Special case of Chain Rule) Let $h(x)$ be a one-variable function which is differentiable at $f(\vec{a})$. Then, $h \circ f$ is differentiable at $\vec{a}$.*

$$\vec{a} \longmapsto f(\vec{a}) \longmapsto (h \circ f)(\vec{a})$$

4. *Any constant function $f(\vec{x}) = c$ is differentiable.*

5. *Coordinate functions $f(\vec{x}) = x_i$ are differentiable.*

Remark. We will discuss general case of chain rule later

Proof of 1,2,3 are similar to those for one variable. (MATH 2050)
The results above give many examples of differentiable functions:

- Polynomials (Sum of products of x_i)
  e.g. $4x^3y^2 + xy^2 - xyz + z^2 \quad (\deg 5)$

- Rational functions (Quotient of polynomials) e.g. $\dfrac{x^3y + z}{x^2 + y^2 + z^2 + 1}$

- If $f(\vec{x})$ is differentiable, then the followings are differentiable:

$$e^{f(\vec{x})}, \quad \sin(f(\vec{x})), \quad \cos(f(\vec{x}))$$

$$\ln(f(\vec{x})) \quad \text{where } f(\vec{x}) > 0$$

$$\sqrt{f(\vec{x})} \quad \text{where } f(\vec{x}) > 0$$
$$\ln|f(\vec{x})| \quad \text{where } f(\vec{x}) \neq 0$$

e.g. $\dfrac{e^{\sqrt{4+\sin(x^2+xy)}}}{\ln\left(1+\cos(x^2y)\right)}$ is differentiable on its domain.

Theorem 7.2. *If a function f is C^1 on an open set $\Omega \subseteq \mathbb{R}^n$, then f is differentiable on Ω.*

Remark. The theorem provides a simple way to verify differentiability if all $\dfrac{\partial f}{\partial x_i}$ can be easily shown to be continuous.

e.g. $f(x, y, z) = xe^{x+y} - \log(x + z)$. The domain of f:

$$\left\{(x, y, z) \in \mathbb{R}^3 : x + z > 0\right\}$$

is open.

$$\frac{\partial f}{\partial x} = e^{x+y} + xe^{x+y} - \frac{1}{x+z}$$
$$\frac{\partial f}{\partial y} = xe^{x+y}$$
$$\frac{\partial f}{\partial z} = -\frac{1}{x+z}$$

Hence, $\dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}, \dfrac{\partial f}{\partial z}$ are all continuous on the open domain of f.

So, f is C^1, and by the theorem is it is differentiable.

Proof of Theorem 7.2. We prove the theorem for the special case where f has two variables.

Let B be an open ball centered at (a, b) such that f_x, f_y are defined on B.

For each fixed $x \in B$, viewing $f(x, y)$ as a one-variable function in y, by the MVT there exists k between b and y such that:

$$f(x, y) - f(x, b) = f_y(x, k)(y - b).$$

Likewise, for fixed $y = b$, there exists h between a and x such that:

$$f(x, b) - f(a, b) = f_x(h, b)(x - a).$$

Hence,

$$f(x, y) - f(a, b) = \underbrace{f(x, y) - f(x, b)}_{f_y(x,k)(y-b)} + \underbrace{f(x, b) - f(a, b)}_{f_x(h,b)(x-a)}$$

We have:

$$\begin{aligned}\left|\frac{\varepsilon(x,y)}{\|(x-a,y-b)\|}\right| &= \left|\frac{f(x,y)-f(a,b)-f_x(a,b)(x-a)-f_y(a,b)(y-b)}{\|(x-a,y-b)\|}\right| \\ &= \left|\frac{[f_y(x,k)-f_y(a,b)](y-b)+[f_x(h,b)-f_x(a,b)](x-a)}{\|(x-a,y-b)\|}\right| \\ &\leq \left|\frac{[f_y(x,k)-f_y(a,b)](y-b)}{\|(x-a,y-b)\|}\right| + \left|\frac{[f_x(h,b)-f_x(a,b)](x-a)}{\|(x-a,y-b)\|}\right| \\ &\leq |f_y(x,k)-f_y(a,b)| + |f_x(h,b)-f_x(a,b)|\end{aligned}$$

Take the limit of both sides of the above inequality as $(x,y) \to (a,b)$. Then, $(x,k),(h,b) \to (a,b)$, and by the continuity of f_x and f_y at (a,b) the right-hand side of the inequality tends to zero.

It follows that:

$$\lim_{(x,y)\to(a,b)} \frac{\varepsilon(x,y)}{\|(x-a,y-b)\|} = 0.$$

So, f is differentiable at (a,b). □

7.2 Gradient and Directional derivative

Definition 7.3. Let $\Omega \subseteq \mathbb{R}^n$ be open , $\vec{a} \in \Omega, f : \Omega \to \mathbb{R}$. The **gradient**, or **gradient vector**, of f at $\vec{a}$ is:

$$\nabla f(\vec{a}) = \left(\frac{\partial f}{\partial x_1}(\vec{a}), \cdots, \frac{\partial f}{\partial x_n}(\vec{a})\right) \in \mathbb{R}^n$$

Example 7.4.

$$\begin{aligned} f(x,y) &= x^2 + 2xy \\ \nabla f(x,y) &= (f_x, f_y) = (2x+2y, 2x) \\ \nabla f(1,2) &= (6,2) \end{aligned}$$

Remark. Using ∇f, the linearization of f at $\vec{a}$ can be expressed as:

$$\begin{aligned} L(\vec{x}) &= f(\vec{a}) + \sum \frac{\partial f}{\partial x_i}(\vec{a})\,(x_i - a_i) \\ &= f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}) \end{aligned}$$

Definition 7.5. Let $\Omega \subseteq \mathbb{R}^n$ be open, $\vec{a} \in \Omega$, $f : \Omega \to \mathbb{R}$.

Let $\vec{u} \in \mathbb{R}^n$ be an unit vector (i.e. $\|\vec{u}\| = 1$) The **directional derivative** of f in the direction of $\vec{u}$ at $\vec{a}$ is:

$$\begin{aligned} D_{\vec{u}} f(\vec{a}) &= \lim_{t\to 0} \frac{f(\vec{a}+t\vec{u}) - f(\vec{a})}{t} \\ &= \text{ the rate of change of } f \text{ in the direction of } \vec{u} \text{ at the point } \vec{a} \end{aligned}$$

Example 7.6. $e_2 = (0,1) \in \mathbb{R}^2$

$$\begin{aligned} D_{e_2} f(a,b) &= \lim_{t\to 0} \frac{f((a,b)+te_2) - f(a,b)}{t} \\ &= \lim_{t\to 0} \frac{f(a,b+t) - f(a,b)}{t} \\ &= \frac{\partial f}{\partial y}(a,b) \end{aligned}$$

Remark. In general, if $e_i = (0,\cdots,0,1,0,\cdots 0) \in \mathbb{R}^n$ (with the i-th entry equal 1), then:

$$D_{e_i} f(\vec{a}) = \frac{\partial f}{\partial x_i}(\vec{a}).$$

Theorem 7.7. *Suppose f is differentiable at $\vec{a}$. Let $\vec{u} \in \mathbb{R}^n$ be a unit vector . Then:*

$$D_{\vec{u}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{u}$$

Remark. Recall that if $\vec{v} \neq \vec{0} \in \mathbb{R}^n$, then the unit vector $\dfrac{\vec{v}}{\|\vec{v}\|}$ is essentially the direction of $\vec{v}$.

Example 7.8. Let $f(x,y) = \arcsin\left(\dfrac{x}{y}\right)$.

Find the rate of change of f at $(1,\sqrt{2})$ in the direction of $\vec{v} = (1,-1)$.

Solution. Let $\vec{u} = \dfrac{\vec{v}}{\|\vec{v}\|} = \left(\dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{2}}\right)$.

Recall: $\dfrac{d}{dz}(\arcsin z) = \dfrac{1}{\sqrt{1-z^2}}$.

Hence.

$$\frac{\partial f}{\partial x} = \frac{1}{\sqrt{1-\left(\frac{x}{y}\right)^2}} \cdot \frac{1}{y},$$

$$\frac{\partial f}{\partial y} = \frac{1}{\sqrt{1-\left(\frac{x}{y}\right)^2}} \cdot \frac{-x}{y^2}.$$

Note that: $f, \dfrac{\partial f}{\partial x}, \dfrac{\partial f}{\partial y}$ are continuous near $(1,\sqrt{2})$.

Hence, f is C^1 near $(1,\sqrt{2})$

So, f is differentiable at $(1,\sqrt{2})$.

By the theorem above, it follows that:

$$\begin{aligned}
D_{\vec{u}}f(1,\sqrt{2}) &= \nabla f(1,\sqrt{2}) \cdot \vec{u} \\
&= \left(\frac{\partial f}{\partial x}(1,\sqrt{2}), \frac{\partial f}{\partial y}(1,\sqrt{2})\right) \cdot \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \\
&= \left(1, -\frac{1}{\sqrt{2}}\right) \cdot \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \\
&= \frac{1}{\sqrt{2}} + \frac{1}{2}
\end{aligned}$$

Proof of Theorem 7.7. Suppose f is differentiable at $\vec{a}$.

Let $L(\vec{x})$ be the linearization of $f(\vec{x})$ at $\vec{a}$.

Then,

$$\begin{aligned}
f(\vec{x}) &= L(\vec{x}) + \varepsilon(\vec{x}) \\
&= f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}) + \varepsilon(\vec{x})
\end{aligned}$$

Put $\vec{x} = \vec{a} + t\vec{u}$:

$$f(\vec{a} + t\vec{u}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (t\vec{u}) + \varepsilon(\vec{a} + t\vec{u})$$

.

Then,

$$\begin{aligned}
D_{\vec{u}}f(\vec{a}) &= \lim_{t\to 0} \frac{f(\vec{a} + t\vec{u}) - f(\vec{a})}{t} \\
&= \lim_{t\to 0} \frac{\nabla f(\vec{a}) \cdot (t\vec{u}) + \varepsilon(\vec{a} + t\vec{u})}{t} \\
&= \nabla f(\vec{a}) \cdot \vec{u} + \lim_{t\to 0} \frac{\varepsilon(\vec{a} + t\vec{u})}{t}
\end{aligned}$$

Differentiability of f at $\vec{a}$ implies that:

$$\lim_{t\to 0} \left| \frac{\varepsilon(a + t\vec{u})}{t} \right| = \lim_{t\to 0} \frac{|\varepsilon(a + t\vec{u})|}{\|(\vec{a} + t\vec{u}) - \vec{a}\|} = 0,$$

It now follows that:

$$D_{\vec{u}}f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{u} + 0 = \nabla f(\vec{a}) \cdot \vec{u}.$$

□

7.3 Geometric Meanings of $\nabla f(\vec{a})$

Suppose f is differentiable at $\vec{a}$ and $\|\vec{u}\| = 1$. Then:

$$D_{\vec{u}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{u}.$$

By the Cauchy-Schwarz Theorem, we have:

$$|\nabla f(\vec{a}) \cdot \vec{u}| \leqslant \|\nabla f(\vec{a})\| \|\vec{u}\| = \|\nabla f(\vec{a})\|$$

Also, if $\nabla f \neq \vec{0}$, then:

$$-\|\nabla f(\vec{a})\| \leq \nabla f(\vec{a}) \cdot \vec{u} \leq \|\nabla f(\vec{a})\|,$$

where each inequality is equality if and only if $\nabla f(\vec{a})$ is parallel to $\vec{u}$.

This means that: f increases (resp. decreases) *most rapidly* in the direction of $\nabla f(\vec{a})$ (resp. $(-\nabla f(\vec{a}))$, at the rate of $\|\nabla f(\vec{a})\|$.

7.4 Properties of the Gradient

Theorem 7.9. *Let $\Omega \subseteq \mathbb{R}^n$ be open. Suppose $f, g : \Omega \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}$ are differentiable, and $c \in \mathbb{R}$ is a constant. Then:*

1. $\nabla(f + g) = \nabla f + \nabla g$
2. $\nabla(cf) = c\nabla f$
3. $\nabla(fg) = g\nabla f + f\nabla g$
4. $\nabla\left(\frac{f}{g}\right) = \frac{g\nabla f - f\nabla g}{g^2}$, *provided that:* $g \neq 0$

Proof of Theorem 7.9. **Exercise**. □

Remark. In the definition of $D_{\vec{u}} f(\vec{a})$, the vector $\vec{u}$ is assumed to be a unit vector. It can also be generalized to $D_{\vec{v}} f(\vec{a})$ for any vector $\vec{v}$ of any length.

In that case,

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

and $D_{\vec{v}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{v}$.

Note that:

$$D_{\vec{v}} f = \begin{cases} \|\vec{v}\| D_{\vec{u}} f & \text{if } \vec{v} \neq \vec{0},\ \vec{u} = \dfrac{\vec{v}}{\|\vec{v}\|} \\ 0 & \text{if } \vec{v} = \vec{0} \end{cases}$$

7.5 Total Differential

(of a real-valued function)

Let $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\vec{a} \in \Omega$.
Consider linearization at $\vec{a}$:

$$f(\vec{x}) = f(\vec{a}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a})\,(x_i - a_i) + \varepsilon(\vec{x})$$

Denote:

$$\Delta f = f(\vec{x}) - f(\vec{a}), \Delta x_i = x_i - a_i$$

Then,

$$\Delta f \approx \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a})\Delta x_i.$$

The approximation is good up to first order, since:

$$\lim_{\vec{x}\to\vec{a}} \frac{\varepsilon(\vec{x})}{\|\vec{x} - \vec{a}\|} = 0.$$

Classically, this first order approximated change is denoted by

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a})dx_i$$

and is called the **total differential** of f at $\vec{a}$.

Example 7.10. Let $V(r, h) = \pi r^2 h$, the volume of a cylinder of radius r and height h.

Observe that V is C^1 on $\mathbb{R}^2$, hence it is differentiable everywhere.
We have:

$$\begin{aligned} dV &= \frac{\partial V}{\partial r}dr + \frac{\partial V}{\partial h}dh \\ &= 2\pi r h dr + \pi r^2 dh \end{aligned}$$

For application:
Suppose we want to approximate change of V when (r, h) changes from $(r, h) = (3, 12)$ to $(3 + 0.08, 12 - 0.3)$
Let

$$dr = \Delta r = 0.08,$$

$$dh = \Delta h = -0.3$$

Then:

$$\begin{aligned}\Delta V \approx dV &\leftarrow \text{ approximated change}\\ &= 2\pi rhdr + \pi r^2 dh\\ &= 2\pi(3)(12)(0.08) + \pi(3)^2(-0.3)\\ &= 3.06\pi \approx 9.61\end{aligned}$$

7.5.1 Properties of the Total Differential

Theorem 7.11. *Suppose $f, g : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ are differentiable and $c \in \mathbb{R}$ is a constant. Then:*

1. $d(f+g) = df + dg$
2. $d(cf) = c\,df$
3. $d(fg) = g\,df + f\,dg$
4. $d\left(\dfrac{f}{g}\right) = \dfrac{g\,df - f\,dg}{g^2}$

Proof of Theorem 7.11. **Exercise.** □

7.6 Summary: Differentiating a real-valued function $f(\vec{x}) = f(x_1, \cdots, x_n)$ at $\vec{a} \in \mathbb{R}^n$

7.6.1 Different types of derivatives

- Directional derivative: $D_{\vec{u}} f(\vec{a}) = \lim_{t\to 0} \dfrac{f(\vec{a}+t\vec{u}) - f(\vec{a})}{t}$ for $\|\vec{u}\| = 1$
- Partial derivative: $\dfrac{\partial f}{\partial x_i}(\vec{a}) = D_{e_i} f(\vec{a}) \quad e_i = (0, \cdots, 0, 1, 0, \cdots, 0)$
- Gradient: $\nabla f(\vec{a}) = \left(\dfrac{\partial f}{\partial x_1}(\vec{a}), \cdots, \dfrac{\partial f}{\partial x_n}(\vec{a})\right)$
- Total differential: $df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{a}) dx_i$
- Higher derivatives: eg $\dfrac{\partial^2 f}{\partial x_1 \partial x_2} = f_{x_2 x_1}$

f is C^k means f and all its partial derivatives up to order k exist and are continuous

7.6.2 Linear Approximation of $f(\vec{x})$ near $\vec{a}$

- $L(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a})$
- $f(\vec{x}) = L(\vec{x}) + \varepsilon(\vec{x})$
- f is differentiable at $\vec{a}$ if $\lim_{\vec{x}\to\vec{a}} \dfrac{\varepsilon(\vec{x})}{\|\vec{x}-\vec{a}\|} = 0 \Rightarrow df \approx \Delta f$

7.6.3 Relations among derivatives

1. $C^\infty \Rightarrow \cdots \Rightarrow C^{k+1} \Rightarrow C^k \Rightarrow \cdots \Rightarrow C^1 \Rightarrow C^0$
2. $$\begin{gathered} f \text{ is } C^1 \text{ on an open set containing } \vec{a} \\ \Downarrow \\ f \text{ is differentiable at } \vec{a}. \\ \Downarrow \\ D_{\vec{u}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{u} \\ \Downarrow \\ D_{\vec{u}} f(\vec{a}) \text{ exists for any unit vector } \vec{u} \in \mathbb{R}^n \\ \Downarrow \\ \frac{\partial f}{\partial x_i}(\vec{a}) \text{ exists for } i = 1, \cdots, n \end{gathered}$$
3. All the $\Rightarrow$ in the reverse direction are false. See next slide for counter examples

Verify the following (counter-) example:

Example 7.12. $f : \mathbb{R} \longrightarrow \mathbb{R}$

$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$

f is differentiable on $\mathbb{R}$ but $f'(x)$ is not continuous at $x = 0$.

Simiarly,

$g(x) = x^{2k-2} f(x)$ is k-time differentiable but $g^{(k)}(x)$ is not continuous at $x = 0$.

Hence, k-time differentiable $\not\Rightarrow C^k$

In particular, $C^{k-1} \not\Rightarrow C^k$

For a multivariable example, let: $h(\vec{x}) = g(x_1)$.

Example 7.13. Let:

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

Then, $D_{\vec{u}}f(0,0)$ exists for any unit vector $\vec{u} \in \mathbb{R}^2$ but f is not continuous at $(0,0)$.

Example 7.14. Let $f(x,y) = |x+y|$.
Then, f is continuous on $\mathbb{R}^2$ but $f_x(0,0)$, $f_y(0,0)$ do not exist.

Example 7.15. Let $f(x,y) = \sqrt{|xy|}$.
Then, $f_x(0,0)$, $f_y(0,0)$ exist, but $D_{\vec{u}}f(0,0)$ does not exist for any $\vec{u} \neq \pm\vec{e}_1, \pm\vec{e}_2$