Chapter 2

Differential Calculus

2.1 Differentiability in one variable

Let $f : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$. We say $f'(a) \in \mathbb{R}$ is the *derivative* of f at a if

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

Note that $f'(a)$ is the slope of the tangent line to the graph of f at point $(a, f(a))$.

We now look at (2.1) from another point of view. Let $m = f'(a)$. From (2.1), we have

$$\lim_{x\to a} \frac{f(x) - f(a) - m(x-a)}{x - a} = \lim_{x\to a} \frac{E(x-a)}{x-a} = 0,$$

where $E(x - a) = f(x) - l(x)$ is the difference between $f(x)$ and its *linear approximation* $l(x)$, here $l(x) = m(x - a) + f(a)$ is the "linear" equation for the tangent line.

Let $h = x - a$, we have $f(a + h) = f(a) + mh + E(h)$, and $E(h)/h \to 0$ as $h \to 0$. This leads to the following definition

Definition 2.1. f is differentiable at a if there is $m \in \mathbb{R}$ such that

$$f(a+h) = f(a) + mh + E(h), \text{ where } \lim_{h\to 0} \frac{E(h)}{h} = 0. \tag{2.2}$$

Note that $m = f'(a)$ is unique when it exists.

Let $S \subset \mathbb{R}$, then f is differentiable on S if it is differentiable at every point of S.

Example 2.2. If $f : \mathbb{R} \to \mathbb{R}$ is a constant function then $f'(x) = 0$ for all $x \in \mathbb{R}$.

If $f(x) = cx$ where c is a fixed number and $x \in \mathbb{R}$, then $f'(x) = c$ for all x.

Remark 2.3. If f is differentiable at a then f is continuous at a.

Proposition 2.4. *Let $a \in \mathbb{R}$ and $f, g : \mathbb{R} \to \mathbb{R}$ be differentiable at a. Then (i) $f \pm g$ are differentiable at a and*

$$(f \pm g)'(a) = f'(a) \pm g'(a). \tag{2.3}$$

(ii) fg is differentiable at a and

$$(fg)'(a) = f'(a)g(a) + f(a)g'(a). \tag{2.4}$$

(iii) If $g(a) \neq 0$, then (f/g) is differentiable at a and

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - g'(a)f(a)}{g^2(a)}. \tag{2.5}$$

In particular,

$$\left(\frac{1}{g}\right)'(a) = -\frac{g'(a)}{g^2(a)}. \tag{2.6}$$

Proof. We prove, for instance (ii). Suppose

$$f(a+h) = f(a) + f'(a)h + E_1(h), \text{ where } \lim_{h\to 0} \frac{E_1(h)}{h} = 0,$$

$$g(a+h) = g(a) + g'(a)h + E_2(h), \text{ where } \lim_{h\to 0} \frac{E_2(h)}{h} = 0.$$

Then $f(a+h)g(a+h) = f(a)g(a) + \{f'(a)g(a) + g'(a)f(a)\}h + E_3(h)$, where

$$E_3(h) = f'(a)g'(a)h^2 + E_1(h)\{g(a) + g'(a)h + E_2(h)\} + E_2(h)\{fa) + f'(a)h\}.$$

Note that
$$\frac{E_3(h)}{h} = f'(a)g'(a)h + \frac{E_1(h)}{h}\{g(a) + g'(a)h + E_2(h)\} + \frac{E_2(h)}{h}\{fa) + f'(a)h\},$$
which goes to zero as $h \to 0$. Therefore (fg) is differentiable at a and its derivative is $(fg)'(a)$ is $f'(a)g(a) + f(a)g'(a)$. ☐

Definition 2.5. Let $S \subset \mathbb{R}^n$, $f : S \to \mathbb{R}$, and $a \in S$.

$f(a)$ is the *maximum* (largest value) of f on S if $f(a) \geq f(x)$ for all $x \in S$.

$f(a)$ is the *minimum* (smallest value) of f on S if $f(a) \leq f(x)$ for all $x \in S$.

f has a *local maximum* at a if there is $r > 0$ such that $f(x) \leq f(a)$ for all $x \in S \cap B(r, a)$.

f has a *local minimum* at a if there is $r > 0$ such that $f(x) \geq f(a)$ for all $x \in S \cap B(r, a)$.

Note that if $f(a)$ is the maximum (respectively, minimum) then it is also a local maximum (respectively, local minimum).

Proposition 2.6. *Suppose f is defined on an open set $I \subset \mathbb{R}$ and $a \in I$. If f has a local maximum or minimum at a and f is differentiable at a then $f'(a) = 0$.*

Proof. Suppose $f(a)$ is a local minimum. Let $\delta > 0$ be such that if $|h| < \delta$, then $a + h \in I$ and $f(a + h) - f(a) \geq 0$. We have
$$f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}.$$
When $0 < h < \delta$, we have $\frac{f(a+h)-f(a)}{h} \geq 0$, letting $h \to 0$ gives $f'(a) \geq 0$. When $-\delta < h < 0$, we have $\frac{f(a+h)-f(a)}{h} \leq 0$, letting $h \to 0$ gives $f'(a) \leq 0$. We conclude $f'(a) = 0$.

☐

Lemma 2.7 (Rolle's theorem)**.** *Suppose $a < b$ and f is differentiable on (a, b) and continuous on $[a, b]$. If $f(a) = f(b)$, then there is $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. Since $[a, b]$ is compact, then there are $x_1, x_2 \in [a, b]$ such that $f(x_1) = M$ is the (absolute) maximum and $f(x_2) = m$ is the (absolute) minimum of f on $[a, b]$.

If $M = m$, then f is a constant function, hence $f'(c) = 0$ for any $c \in (a, b)$.

If $M \neq m$, then $M \neq L = f(a) = f(b)$ or $m \neq L$. Suppose $M \neq L$ then $c = x_1 \neq a, b$, hence $c \in (a, b)$. Since f is differentiable on the open interval (a, b) and has a local maximum at $c \in (a, b)$, then by Proposition 2.6 we have $f'(c) = 0$. □

Theorem 2.8 (Mean value theorem I)**.** *Suppose f is continuous on $[a, b]$ and is differentiable on (a, b). Then there is a point $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{2.7}$$

Note that $\frac{f(b)-f(a)}{b-a}$ is the slope of the straight line going through $(a, f(a))$ and $(b, f(b))$.

Proof. Let

$$g(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a) - f(x).$$

Then g is continuous on $[a, b]$ and is differentiable on (a, b). Note that $g(a) = g(b) = 0$ and $g'(x) = \frac{f(b)-f(a)}{b-a} - f'(x)$. By Rolle's lemma, there is $c \in (a, b)$ such that $g'(c) = 0$, hence we obtain (2.7). □

Theorem 2.9. *Suppose f is differentiable on an open interval I. (a) If $|f'(x)| \leq C$ for all $x \in I$ then $|f(b) - f(a)| \leq C|b - a|$ for all $a, b \in I$.*
(b) If $f'(x) = 0$ for all $x \in I$ then f is constant in I.
(c) If $|f'(x)| \geq 0$ (resp., $> 0, \leq, < 0$) for all $x \in I$ then f is increasing (resp., strictly increasing, decreasing, strictly decreasing) on I.

Proof. Let $a, b \in I$ and $a < b$, then f continuous on $[a, b]$ and is differentiable on (a, b). By the Mean Value Theorem 2.8, there is $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a).$$

We easily prove (a)–(c). For example, if $f'(x) < 0$ for all $x \in I$ then $f'(c) < 0$, therefore $f(b) - f(a) < 0$ for any $b > a$; that means f is strictly decreasing in I. □

Theorem 2.10 (Mean value theorem II)**.** *Suppose f and g are continuous on $[a, b]$ and is differentiable on (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$. Then there is a point $c \in (a, b)$ such that*

$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}. \tag{2.8}$$

Proof. Apply Rolle's lemma for the following function

$$h(x) = [f(x) - f(a)][g(b) - g(a)] - [g(x) - g(a)][f(b) - f(a)].$$

□

Definition 2.11. We have the following notion of limits

- Let $f : (d, a) \to \mathbb{R}^m$ and $L \in \mathbb{R}^m$. Then $\lim_{x\to a-} f(x) = L$ if

$$\forall \varepsilon > 0, \exists \delta > 0, \forall x \in (d, a) : a - \delta < x < a \implies |f(x) - L| < \varepsilon. \tag{2.9}$$

- Let $f : (a, b) \to \mathbb{R}^m$ and $L \in \mathbb{R}$. Then $\lim_{x\to a+} f(x) = L$ if

$$\forall \varepsilon > 0, \exists \delta > 0, \forall x \in (a, b) : a < x < a + \delta \implies |f(x) - L| < \varepsilon. \tag{2.10}$$

- Let $f : (c, \infty) \to \mathbb{R}^m$ and $L \in \mathbb{R}$. Then $\lim_{x\to\infty} f(x) = L$ if

$$\forall \varepsilon > 0, \exists M > 0, \forall x \in (c, \infty) : x > M \implies |f(x) - L| < \varepsilon. \tag{2.11}$$

- Let $f : (-\infty, c) \to \mathbb{R}^m$ and $L \in \mathbb{R}$. Then $\lim_{x\to-\infty} f(x) = L$ if

$$\forall \varepsilon > 0, \exists M > 0, \forall x \in (-\infty, c) : x < -M \implies |f(x) - L| < \varepsilon. \tag{2.12}$$

- Let $f : \mathbb{R}^n \to \mathbb{R}$, $a \in \mathbb{R}^n$. Then $\lim_{x\to a} f(x) = \infty$ if

$$\forall M > 0, \exists \delta > 0, \forall x \in \mathbb{R}^n : 0 < |x - a| < \delta \implies f(x) > M. \tag{2.13}$$

- Let $f : \mathbb{R}^n \to \mathbb{R}$, $a \in \mathbb{R}^n$. Then $\lim_{x\to a} f(x) = -\infty$ if

$$\forall M > 0, \exists \delta > 0, \forall x \in \mathbb{R}^n : 0 < |x - a| < \delta \implies f(x) < -M. \quad (2.14)$$

Note that if $f : (d, a) \cup (a, b) \to \mathbb{R}^m$ and $L \in \mathbb{R}^m$ then

$$\lim_{x\to a} f(x) = L \iff \lim_{x\to a-} f(x) = \lim_{x\to a+} f(x) = L. \quad (2.15)$$

Theorem 2.12 (L'Hôpital's rule I)**.** *Suppose f and g are differentiable on (a, b) and*

$$\lim_{x\to a+} f(x) = \lim_{x\to a+} g(x) = 0. \quad (2.16)$$

If g' never vanishes on (a, b) and

$$\lim_{x\to a+} \frac{f'(x)}{g'(x)} = L, \quad (2.17)$$

then

$$\lim_{x\to a+} \frac{f(x)}{g(x)} = L. \quad (2.18)$$

Proof. Extend $f(a) = 0, g(a) = 0$. For $x \in (a, b)$, we have f, g are continuous on $[a, x]$ and differentiable on (a, x). By Theorem 2.10, there is $c \in (a, x)$ such that

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f'(c)}{g'(c)}.$$

Note that $c \to a+$ and $x \to a+$. Letting $x \to a+$ and using (2.17), we obtain (2.18). □

Remark 2.13. The theorem still holds if we replace $\lim_{x\to a+}$ by $\lim_{x\to a-}$, $\lim_{x\to a}$, $\lim_{x\to\infty}$, $\lim_{x\to-\infty}$ and the domains of f, g are appropriate.

Theorem 2.14 (L'Hôpital's rule II)**.** *Suppose f and g are differentiable on (a, b) and*

$$\lim_{x\to a+} |f(x)| = \lim_{x\to a+} |g(x)| = \infty. \quad (2.19)$$

If g' never vanishes on (a, b) and

$$\lim_{x\to a+} \frac{f'(x)}{g'(x)} = L, \quad (2.20)$$

then

$$\lim_{x\to a+}\frac{f(x)}{g(x)} = L. \tag{2.21}$$

Theorem 2.15 (Chain rule)**.** *Let $f, g : \mathbb{R} \to \mathbb{R}$ and $a \in \mathbb{R}$. Let $g(a) = b$ and suppose that g is differentiable at a, and f is differentiable at b. Then $f \circ g$ is differentiable at a and*

$$(f \circ g)'(a) = f'(b)g'(a). \tag{2.22}$$

Proof. We have

$$g(a + h) = g(a) + g'(a)h + E_1(h), \text{ where } \lim_{h\to 0}\frac{E_1(h)}{h} = 0,$$

$$f(b + h) = f(b) + f'(b)h + E_2(k), \text{ where } \lim_{k\to 0}\frac{E_2(k)}{k} = 0.$$

Then $(f\circ g)(a+h) = f(g(a+h)) = f(b+k)$ where $k = k(h) = g'(a)h+E_1(h)$. We have

$$\begin{aligned}(f \circ g)(a + h) &= f(b) + f'(b)\{g'(a)h + E_1(h)\} + E_2(k(h))\\ &= (f \circ g)(a) + f'(b)g'(a)h + E_3(h),\end{aligned} \tag{2.23}$$

where $E_3(h) = f'(b)E_1(h) + E_2(k(h))$. Note that

$$\frac{E_3(h)}{h} = f'(b)\frac{E_1(h)}{h} + \frac{E_2(k(h))}{h}$$

Claim: $\lim_{h\to 0}\frac{E_2(k(h))}{h} = 0$.

Suppose the claim is true, then $\lim_{h\to 0} E_3(h)/h = 0$. Hence, according to the Definition 2.1, we infer from (2.23) that $f \circ g$ is differentiable at a and (2.22).

Proof of the claim: The idea is that

$$\frac{E_2(k(h))}{h} = \frac{E_2(k(h))}{k(h)}\frac{k(h)}{h}.$$

Since

$$\lim_{h\to 0} k(h) = 0, \quad \lim_{k\to 0}\frac{E_2(k)}{k} = 0, \quad \text{and } \lim_{h\to 0}\frac{k(h)}{h} = g'(a),$$

we obtain $\lim_{h\to 0} \frac{E_3(h)}{h} = 0$. This argument can be easily made rigorous (to take care of the case $k(h) = 0$). However, the direct proof can go as follows:

Let $M = |g'(a)| + 1$. Since $\lim_{h\to 0} \frac{k(h)}{h} = g'(a)$, there is $\delta_1 > 0$ such that $|k(h)| \leq M|h|$ for $0 < |h| < \delta_1$.

Let $\varepsilon > 0$. Since $\lim_{k\to 0} \frac{E_2(k)}{k} = 0$, there is $\delta_2 > 0$ such that $|E_2(k)| \leq (\varepsilon/M)|k|$ for $|k| < \delta_2$ (note that $E_2(0) = 0$). Let $\delta = \min\{\delta_1, \delta_2/M\}$, then for $0 < |h| < \delta$, we have $|k(h)| \leq M|h| \leq \delta_2$ and hence

$$|E_2(k(h))| \leq (\varepsilon/M)|k(h)| \leq (\varepsilon/M)M|h| = \varepsilon|h|.$$

Therefore $\lim_{h\to 0} E_2(k(h))/h = 0$. □

Differentiability of vector-valued functions. Let $f = (f_1, f_2, \ldots, f_m) : \mathbb{R} \to \mathbb{R}^m$ be a vector-valued function, where $f_j : \mathbb{R} \to \mathbb{R}$, for $j = 1, 2, \ldots m$. Let $a \in \mathbb{R}$. Then the derivative of f at a is the vector

$$f'(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = (f_1'(a), f_2'(a), \ldots, f_n'(a)). \tag{2.24}$$

whenever the involved quantities are defined. If $f'(a)$ exists then we say f is differentiable at a. In fact, $f'(a)$ is the unique vector $v \in \mathbb{R}^m$ such that

$$f(a+h) = f(a) + hv + E(h), \text{ where } E(h) \in \mathbb{R}^m, \quad \lim_{h\to 0} \frac{E(h)}{h} = 0. \tag{2.25}$$

Curves and tangent vectors. See text, p.50.

Higher order derivatives. Just as in lower calculus course.

2.2 Differentiability in several variables

2.2.1 Real-valued functions

Partial derivatives. Let $f : \mathbb{R}^n \to \mathbb{R}$, $a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. Partial derivative of f with respect to variable x_j at a is

$$\frac{\partial f}{\partial x_j}(a) = \lim_{h \to 0} \frac{f(a_1, \ldots, a_{j-1}, a_j + h, a_{j+1}, \ldots, a_n) - f(a_1, \ldots, a_j, \ldots, a_n)}{h} \tag{2.26}$$

Other notation: $f_{x_j}, \partial_j f, \partial_{x_j} f$.

Gradient vector and Differentiability. Let $S \subset \mathbb{R}^n$ be open, $f : S \to \mathbb{R}$, $a \in S$. We say f is differentiable at a if there is $c \in \mathbb{R}^n$ such that

$$f(a+h) = f(a) + c \cdot h + E(h), \text{ where } \lim_{h \to 0} \frac{E(h)}{|h|} = 0. \tag{2.27}$$

The vector c is the gradient of f at a and is denoted by $\nabla f(a)$.

Tangent planes. For $n = 2$, $f = f(x) = f(x_1, x_2)$ the graph of $z = f(x)$ is a surface in $\mathbb{R}^3$. Let $P = (a, f(a))$ be a point on the surface. The equation for the tangent plane of the surface at P is:

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

Theorem 2.16 (Chain Rule)**.** *Let $g(t) = (g_1, g_2, \ldots, g_n) : \mathbb{R}^m \to \mathbb{R}^n$, $f(x) : \mathbb{R}^n \to \mathbb{R}$, $a \in \mathbb{R}^m, b = g(a) \in \mathbb{R}^n$. If g is differentiable at a and f is differentiable at b then $f \circ g$ is differentiable at a and*

$$\frac{\partial (f \circ g)}{\partial t_k}(a) = \frac{\partial f}{\partial x_1}(b)\frac{\partial g_1}{\partial t_k}(a) + \frac{\partial f}{\partial x_2}(b)\frac{\partial g_2}{\partial t_k}(a) + \ldots + \frac{\partial f}{\partial x_n}(b)\frac{\partial g_n}{\partial t_k}(a), \tag{2.28}$$

for $k = 1, 2, \ldots, m$. Briefly, we have

$$\frac{\partial (f \circ g)}{\partial t_k}(a) = \nabla f(b) \cdot \frac{\partial g}{\partial t_k}(a), \tag{2.29}$$

for $k = 1, 2, \ldots, m$.

Directional derivatives. Let $u \in \mathbb{R}^n, |u| = 1$, then

$$\partial_u f(a) = \lim_{h \to 0} \frac{f(a + hu) - f(a)}{h}. \tag{2.30}$$

We have

$$\partial_u f(a) = \nabla f(a) \cdot u. \tag{2.31}$$

By Cauchy-Schwarz's inequality $|\partial_u f(a)| \leq |\nabla f(a)||u| = |\nabla f(a)|$. Hence $\partial_u f(a)$ attains its maximum value $|\nabla f(a)|$ when $u = \lambda \nabla f(a)$ for some $\lambda > 0$.

2.2.2 Vector-valued functions

Definition 2.17. Let $f : \mathbb{R}^n \to \mathbb{R}^m$, $a \in \mathbb{R}^n$. We say f is differentiable at a if there is a $m \times n$ matrix L such that

$$f(a + h) = f(a) + Lh + E(h), \text{ where } E(h) \in \mathbb{R}^m, \quad \lim_{h \to 0} \frac{E(h)}{|h|} = 0. \tag{2.32}$$

The matrix L, denoted by $Df(a)$ (or $f'(a)$), is called the (Fréchet) derivative of f at a.

Proposition 2.18. *If $Df(a)$ exists, then it is unique.*

Proposition 2.19. *If f is differentiable at a then f is continuous at a.*

Proposition 2.20. *Let $f = (f_1, f_2, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}^n$. Then the partial derivatives $\partial_{x_j} f_i(a)$, for $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, exist and the matrix $Df(a)$ is*

$$Df = \left(\frac{\partial f_i}{\partial x_j} \right)_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} = \begin{pmatrix} Df_1 \\ Df_2 \\ \vdots \\ Df_m \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}. \tag{2.33}$$

Theorem 2.21 (Chain Rule)**.** *Suppose $g : \mathbb{R}^k \to \mathbb{R}^n$ is differentiable at $a \in \mathbb{R}^k$ and $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $b = g(a) \in \mathbb{R}^n$. Then their composition $H = f \circ g : \mathbb{R}^k \to \mathbb{R}^m$ is differentiable at a, and*

$$DH(a) = DF(b)Dg(a). \tag{2.34}$$

Note that Df is an $m \times n$ matrix, Dg is an $n \times k$ matrix and DH is an $m \times k$ matrix.

Theorem 2.22. *Let $S \subset \mathbb{R}^n$ be open, $f : S \to \mathbb{R}$, and $a \in S$. Suppose all partial derivatives $\partial_j f(a)$, for $j = 1, 2, \ldots, n$, exist in a neighborhood of a and are continuous at a, then f is differentiable at a.*

2.3 The Mean Value Theorem

The following notation is not standard and is only used in this lecture note.

Let $a, b \in \mathbb{R}^n$, we denote the line segments whose endpoints are a and b by

$$[a, b] = \{(1-t)a + tb \ : \ t \in [0, 1]\},$$

and

$$(a, b) = \{(1-t)a + tb \ : \ t \in (0, 1)\},$$

Note that $l(t) = (1-t)a + tb$, for $t \in [0, 1]$, is the equation for the closed line segment $[a, b]$, and $l(0) = a$, $l(1) = b$.

A subset S of $\mathbb{R}^n$ is called *convex* if for any $a, b \in S$, we have $[a, b] \subset S$. Note that every convex set is connected.

Theorem 2.23. *Let S be an open subset of $\mathbb{R}^n$ and $a, b \in S$ such that $[a, b] \subset S$. Suppose $f : S \to \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b), then there is a point $c \in [a, b]$ such that*

$$f(b) - f(a) = \nabla f(c) \cdot (b - a).$$

Corollary 2.24. *Suppose f is differentiable on an open convex set $S \subset \mathbb{R}^n$ and $|\nabla f(x)| \leq M$ for all $x \in S$. Then $|f(b) - f(a)| \leq M|b - a|$ for all $a, b \in S$.*

Remark: We can use this to prove the uniform continuity of a function.

Corollary 2.25. *If S is convex, f is differentiable on S and $\nabla f(x) = 0$ for all $x \in S$, then f is constant on S.*

Corollary 2.25 still holds true when S is only connected.

Theorem 2.26. *Suppose f is differentiable on an open connected set $S \subset \mathbb{R}^n$ and $\nabla f(x) = 0$ for all $x \in S$. Then f is constant on S.*

2.4 Higher-order partial derivatives

See Section 2.6 of the textbook.

Suppose f is defined on an open set $S \subset \mathbb{R}^n$ and $\partial_{x_j} f$, for some $j \in \{1, 2, \ldots, n\}$, exists on S. Then whenever it makes sense, we have the second-order derivative $\partial_{x_i}\left[\partial_{x_j} f\right]$.

Notation:

$$\frac{\partial^2 f}{\partial x_i \partial x_j},\ f_{x_j x_i},\ f_{ji},\ \partial_{x_i}\partial_{x_j} f,\ \partial_i \partial_j f.$$

In particular,

$$\frac{\partial^2 f}{\partial x_j^2},\ f_{x_j x_j},\ f_{jj},\ \partial_{x_j}^2 f,\ \partial_j^2 f.$$

Similarly, we may have third-order partial derivatives $\partial_{x_k}\partial_{x_i}\partial_{x_j} f$ where $j, i, k \in \{1, 2, \ldots, n\}$; or the k-order partial derivatives

$$\partial_{x_{j_k}} \ldots \partial_{x_{j_2}} \partial_{x_{j_1}} f,$$

for $k \in \mathbb{N}$ and $j_1, j_2, \ldots, j_k \in \{1, 2, \ldots, n\}$.

For our convention, the zero-order derivative of f is just f itself.

Definition 2.27. Let $U \subset \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$.

The function f is said to be of class C^k on U if all the partial derivatives of f *up to* order k exist and are continuous on U. Notation $f \in C^k(U)$.

If all partial derivatives of f of all orders exist and are continuous on U then f is said of class C^∞. Notation $f \in C^\infty(U)$.

In the case of vector-valued functions, $f = (f_1, f_2, \ldots, f_m)$ is said of class C^k, (or C^∞,) if each f_j, for $j = 1, 2, \ldots, m$, is of class C^k, (or C^∞).

Theorem 2.28. *Let f be a function defined in an open set $S \subset \mathbb{R}^n$. Suppose $a \in S$ and $i, j \in \{1, 2, \ldots, n\}$. If the derivatives $\partial_i f$, $\partial_j f$, $\partial_i \partial_j f$ and $\partial_j \partial_i f$ exist in S and are continuous at a, then $\partial_i \partial_j f(a) = \partial_j \partial_i f(a)$.*

Corollary 2.29. *If $f \in C^2(S)$ where $S \subset \mathbb{R}^n$ is open, then $\partial_i \partial_j f = \partial_j \partial_i f$ on S for all i, j.*

For higher order derivatives, we have the following theorem

Theorem 2.30. *If $f \in C^k(S)$ where $S \subset \mathbb{R}^n$ is open, then*

$$\partial_{i_1} \partial_{i_2} \ldots \partial_{i_k} f = \partial_{j_1} \partial_{j_2} \ldots \partial_{j_k} f,$$

whenever the sequence $\{j_1, j_2, \ldots, j_k\}$ is a reordering of $\{i_1, i_2, \ldots, i_k\}$.

Multi-index Notation. A multi-index is an n-tuple of non-negative integers:

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n), \quad \alpha_j \in \{0, 1, 2, \ldots\}.$$

Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ be a multi-index, $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$. We define

$$|\alpha| = \alpha_1 + \alpha_2 + \ldots + \alpha_n, \quad \alpha! = \alpha_1! \alpha_2! \ldots \alpha_n!,$$

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n},$$

$$\partial^\alpha f = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \ldots \partial_n^{\alpha_n} f = \frac{\partial^{|\alpha|} f}{\partial {x_1}^{\alpha_1} \partial {x_2}^{\alpha_2} \ldots \partial {x_n}^{\alpha_n}}.$$

Recall $0! = 1,\ 1! = 1,\ 2! = 2(1!) = 2,\ k! = k[(k-1)!] = 1 \cdot 2 \cdot \ldots \cdot k$.

The number $|\alpha|$ is called the *order* or *degree* of α. Also, $|\alpha|$ is the order of the partial derivative $\partial^\alpha f$.

Theorem 2.31 (Multinomial Theorem)**.** *For any $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $k \in \mathbb{N}$, we have*

$$(x_1 + x_2 + \ldots + x_n)^k = \sum_{|\alpha| = k} \frac{k!}{\alpha!} x^\alpha.$$

Particularly, when $n = 2$,

$$(x_1 + x_2)^k = \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} x^j.$$

2.5 Taylor's Theorem

We only present Taylor's theorem with Lagrange's remainder.

2.5.1 In one variable

We aim to approximate the value of a function f near a using the polynomials. The following was explained in details in class.

We write $f(a+h) = P_{a,k}(h) + R_{a,k}(h)$, where $P_{a,k}(h)$ is the k-order Taylor polynomial

$$P_{a,k}(h) = f(a) + f'(a)h + \frac{f''(a)}{2}h^2 + \ldots + \frac{f^{(k)}(a)}{k!}h^k = \sum_{j=0}^{k} \frac{f^{(j)}(a)}{j!}h^j.$$

We expect to have

$$\lim_{h \to 0} \frac{R_{a,k}(h)}{h^k} = 0.$$

Theorem 2.32. *Suppose f is $k+1$ times differentiable on an interval $I \subset \mathbb{R}$ and $a \in I$. For each $h \in \mathbb{R}$ such that $a + h \in I$, there is a point c between 0 and h such that*

$$R_{a,k}(h) = \frac{f^{(k+1)}(a+c)}{(k+1)!}h^{k+1}.$$

The proof of the above theorem requires a generalization of Rolle's Lemma for higher derivatives (see Lemma 2.62 in the text).

Corollary 2.33. *If $|f^{(k+1)}(x)| \leq M$ for all $x \in I$ then*

$$\lim_{h \to 0} \frac{R_{a,k}(h)}{h^k} = 0.$$

See Proposition 2.65 in the text for some examples of Taylor polynomials.

2.5.2 In several variables

Theorem 2.34. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is of class C^{k+1} on an open convex set S. If $a, a+h \in S$, then $f(a+h) = P_{a,k}(h)R_{a,k(h)}$ where*

$$P_{a,k}(h) = \sum_{|\alpha|\leq k} \frac{\partial^\alpha f(a)}{\alpha!} h^\alpha,$$

$$R_{a,k}(h) = \sum_{|\alpha|=k+1} \frac{\partial^\alpha f(a+ch)}{\alpha!} h^\alpha,$$

for some $c \in (0,1)$.

Corollary 2.35. *If, in addition to Theorem 2.34, we have $|\partial^\alpha f(x)| \leq M$ for all $x \in S$ and $|\alpha| = k+1$, then*

$$|R_{a,k}(h)| \leq \frac{M}{(k+1)!}(|h_1| + |h_2| + \ldots + |h_n|)^{k+1},$$

and consequently,

$$\lim_{h\to 0} \frac{R_{a,k}(h)}{|h|^k} = 0.$$

2.6 Critical Points

Theorem 2.36. *Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$. If f has a local maximum or local minimum at $a \in S$ and f is differentiable at a, then $\nabla f(a) = 0$.*