12 Differentiability in one variable

Definition 12.1. (*differentiability*)
Let $f:[a,b]\to\mathbb{R}$ be a function and $x\in[a,b]$.
The function f' is defined by

$$f'(x)=\begin{cases}\displaystyle\lim_{t\to x}\frac{f(t)-f(x)}{t-x}, & \text{if } x\in(a,b),\\ \displaystyle\lim_{t\to x^+}\frac{f(t)-f(x)}{t-x}, & \text{if } x=a,\\ \displaystyle\lim_{t\to x^-}\frac{f(t)-f(x)}{t-x}, & \text{if } x=b,\end{cases}\qquad \text{provided the limit exists.}$$

More precisely, the latter two are often denoted by $f'(a^+)$ and $f'(b^-)$ respectively.
The domain of f' is the set of points x at which the limit exists.
*The function f' is called the **derivative of** f.*
*If f' is defined at x, we say that f is **differentiable at** x.*
If f is differentiable at every point of a set $E\subset[a,b]$,
*we say that f is **differentiable on** E.*

Definition 12.2. (*higher order derivatives*)
If f has a derivative f' on an interval, and f' is differentiable,
*we denote the derivative of f' by f'' or $f^{(2)}$ and call f'' the **second derivative of** f.*
By the same way, we can define functions $f^{(2)},f^{(3)},\cdots,f^{(n-1)},f^{(n)}$
by each of which is the derivative of the previous one.
*We call $f^{(n)}$ the n**-th derivative of** f.*

Theorem 12.3. (*little o*)
Let $f:[a,b]\to\mathbb{R}$ be a function and $x\in[a,b]$.
Then f is differentiable at $x\in(a,b)$, $x=a$, or $x=b$
$\Longleftrightarrow$ *there exists $m\in\mathbb{R}$ such that $f(x+h)=f(x)+mh+o(h)$ for all h close to $0,0^+$, or 0^-, in **Landau notation**,*
*where $o(h)$, called **little** o, is a real-valued function of h which satisfies $\displaystyle\lim_{h\to0,0^+,or\ 0^-}\frac{o(h)}{h}=0$, respectively;*
in this case, $f'(x)=m$.

Theorem 12.4. (*differentiability implies continuity*)
Let $f:[a,b]\to\mathbb{R}$ be a function.
If f is differentiable at $x\in[a,b]$, then f is continuous at x.

Theorem 12.5. (*basic rules of differentiability*)
Let $f,g:[a,b]\to\mathbb{R}$ be functions and let $x\in[a,b]$.
If f,g are differentiable at x, then $f+g$, fg, and $\dfrac{f}{g}$ are differentiable at x, and

1. ***the sum and difference rules:*** $(f\pm g)'(x)=f'(x)\pm g'(x)$;
2. ***the product rule:*** $(fg)'(x)=f'(x)g(x)+f(x)g'(x)$; *and*
3. ***the quotient rule:*** $\left(\dfrac{f}{g}\right)'(x)=\dfrac{f'(x)g(x)-f(x)g'(x)}{g^2(x)}$ *provided $g(x)\neq0$.*

Example 12.6. (*basic differentiable functions*)
Every polynomial is differentiable and so is every rational function,
except at the points where the denominator is zero.

Theorem 12.7 (chain rule)**.**
Let $g:[a,b]\to\mathbb{R}$ be a function, let I be an interval with $I\supset g([a,b])$,
and let $f:I\to\mathbb{R}$ be a function.
If g is differentiable at $x\in[a,b]$ and f is differentiable at $g(x)$,
then $f\circ g$ is differentiable at x and $(f\circ g)'(x)=f'(g(x))g'(x)$.

Proposition 12.8. (*for exercise*)
Let (a,b) be an interval in $\mathbb{R}$ and $f:(a,b)\to\mathbb{R}$ be a differentiable function.

1. *If $\displaystyle\lim_{x\to a^+}f'(x)$ exists, as a finite number, and $a\in\mathbb{R}$, then $\displaystyle\lim_{x\to a^+}f(x)$ exists.*
2. *If, in addition to item 1, f is continuous on $[a,b)$, then f is differentiable at a and $f'(a^+)=\displaystyle\lim_{x\to a^+}f'(x)$.*
3. *If, in addition to item 1, f is continuous on $(a,b]$, then f is uniformly continuous on $(a,b]$.*
4. *In item 1, the condition $a\in\mathbb{R}$ is necessary, see the example $f(x)=\sin\sqrt{x}$ on $(-\infty,0)$.*
   In item 3, the function might have unbounded derivative, for instance $f(x)=\sqrt{-x}$ on $(-1,0]$.

13 The generalized MVT, Rolle's theorem and the monotonicity theorem

Definition 13.1. (*local extreme*)
Let (X,d) be a metric space and $f : X \to \mathbb{R}$ be a function.
*We say that f has a **local maximum** (resp. **local minimum**) at $p \in X$*
if there exists $\delta > 0$ such that
if $x \in X$ with $d(x,p) < \delta$ then $f(x) \le f(p)$ (resp. $f(x) \ge f(p)$).

Theorem 13.2 (the critical point theorem in one variable)**.**
Let $f : [a,b] \to \mathbb{R}$ be a function, where $a < b$.
If f is differentiable at $p \in (a,b)$, and f has a local maximum or a local minimum at p,
then $f'(p) = 0$.

Theorem 13.3 (Rolle's theorem)**.**
Let $f : [a,b] \to \mathbb{R}$ be function, where $a < b$.

1. *If f is continuous on $[a,b]$ and differentiable on (a,b), and $f(a) = f(b)$*
   then there exists a point $c \in (a,b)$ such that $f'(c) = 0$.
2. *If $f^{(n)}$ is differentiable on $[a,b]$, $f(a) = f(b)$, and $f^{(i)}(a) = 0$ for all $1 \le i \le n$,*
   then there exists a point $c \in (a,b)$ such that $f^{(n+1)}(c) = 0$.

Theorem 13.4 (the generalized mean value theorem)**.**
Let $f, g : [a,b] \to \mathbb{R}$ be functions, where $a < b$.
If f and g are continuous on $[a,b]$ and differentiable on (a,b),
then there exists a point $c \in (a,b)$ such that $(f(b) - f(a))\, g'(c) = f'(c)\, (g(b) - g(a))$.
*The special case when $g(x) = x$ is referred to as the **mean value theorem**.*

Definition 13.5. (*monotonicity of functions*)
Let $f : A \subset \mathbb{R} \to \mathbb{R}$ be a function.

1. *We say that f is **increasing** (resp. **strictly increasing**) on A*
   if $f(x_1) \le f(x_2)$ (resp. $f(x_1) < f(x_2)$) for all $x_1, x_2 \in A$ with $x_1 < x_2$.
2. *We say that f is **decreasing** (resp. **strictly decreasing**) on A*
   if $f(x_1) \ge f(x_2)$ (resp. $f(x_1) > f(x_2)$) for all $x_1, x_2 \in A$ with $x_1 < x_2$.
3. *We say that f is **monotonic** (resp. **strictly monotonic**) on A*
   if f is either increasing or decreasing
   (resp. either strictly increasing or strictly decreasing) on A.

Theorem 13.6 (the monotonicity theorem)**.**
Let $f : (a,b) \to \mathbb{R}$ be a differentiable function.

1. *If there exists $M > 0$ such that $|f'(x)| \le M$ for all $x \in (a,b)$,*
   then $|f(x_1) - f(x_2)| \le M|x_1 - x_2|$ for all $x_1, x_2 \in (a,b)$.
2. *If $f'(x) = 0$ for all $x \in (a,b)$, then f is constant on (a,b).*
3. *If $f'(x) \ge 0$ (resp. $f'(x) > 0$) for all $x \in (a,b)$,*
   then f is increasing (resp. strictly increasing) on (a,b).
4. *If $f'(x) \le 0$ (resp. $f'(x) < 0$) for all $x \in (a,b)$,*
   then f is decreasing (resp. strictly decreasing) on (a,b).
5. *If $f'(x) > 0$ for all $x \in (a,b)$ except finitely many points at which $f'(x) = 0$,*
   then f is strictly increasing on (a,b). (exercise)

Proposition 13.7.
There exists a non-constant function from $\mathbb{R}$ to $\mathbb{R}$
such that $f^{(n)}(0) = 0$ and $f^{(n)}(x)$ exists for all $x \in \mathbb{R}$ and for all $n \in \mathbb{N}$. (exercise)

14 L'Hôpital's rule in one variable and Darboux's theorem

Theorem 14.1 (L'Hôpital's rule).
Let $f, g : (a,b) \to \mathbb{R}$ be differentiable on $(a,b)\setminus\{c\}$ satisfying that $g'(x) \neq 0$ for all $x \in (a,b)\setminus\{c\}$, where $-\infty \leq a < b \leq +\infty$ and $a \leq c \leq b$; here we use the ordering $<$ on the extended real line.
If $\dfrac{f(x)}{g(x)}$ has the ***indeterminate form*** *$\dfrac{0}{0}$ or $\dfrac{\infty}{\infty}$ at $x = c^+$,*
i.e., $\lim\limits_{x\to c^+} f(x) = \lim\limits_{x\to c^+} g(x) = 0$ or $\lim\limits_{x\to c^+} g(x) = \pm\infty$ (in this case, we do not need to control $\lim\limits_{x\to c^+} f(x)$),
then $\lim\limits_{x\to c^+} \dfrac{f(x)}{g(x)} = \lim\limits_{x\to c^+} \dfrac{f'(x)}{g'(x)}$, provided $\lim\limits_{x\to c^+} \dfrac{f'(x)}{g'(x)}$ exists or is $\pm\infty$.
The analogous statement is true if $x \to c^-$, $x \to +\infty$, or $x \to -\infty$.

Corollary 14.2.
For any $a, b > 0$ we have

$$\lim_{x\to+\infty} \frac{x^a}{e^{bx}} = \lim_{x\to+\infty} \frac{(\ln x)^b}{x^a} = \lim_{x\to 0^+} \frac{(\ln x)^b}{x^{-a}} = 0 \text{ and } \lim_{x\to 0}(1 \pm ax)^{\frac{\pm b}{x}} = e^{sgn((\pm a)(\pm b))ab} \text{ (exercise).}$$

Remark 14.3.
There is a function f on $\mathbb{R}$ such that f is differentiable on $\mathbb{R}$ but f' is not continuous at some point. (exercise)

Theorem 14.4 (Darboux's theorem *or* intermediate value property for derivatives).
Let $f : [a,b] \to \mathbb{R}$ be a differentiable function, where $a < b$.
(Notice that f' might not be continuous on $[a,b]$.)
If $s \in \mathbb{R}$ with $f'(a) < s < f'(b)$ or $f'(a) > s > f'(b)$,
then there exists $c \in (a,b)$ such that $f'(c) = s$. (exercise)

Definition 14.5 (little o and big O).
Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a function and $g : \mathbb{R}^m \to \mathbb{R}^k$ be a function, and let $p \in \mathbb{R}^m$.
We say that f is ***little*** *o* ***of*** *$g(x)$ at p, and write $f(x) = o(g(x))$ as $x \to p$,*
if for any $M > 0$ there exists $\delta > 0$ such that $\|f(x)\| \leq M\|g(x)\|$ for all x with $\|x - p\| < \delta$,
(equivalently, $\lim_{x\to p} \frac{f(x)}{\|g(x)\|} = 0$ provided that $g(x) \neq 0$).
We say that f is ***big*** *O* ***of*** *$g(x)$ at p, and write $f(x) = O(g(x))$ as $x \to p$,*
if there exist $M > 0$ and $\delta > 0$ such that $\|f(x)\| \leq M\|g(x)\|$ for all x with $\|x - p\| < \delta$.
If $m = 1$, then the similar definitions work for $p = \pm\infty$.

15 Differentiability of vector-valued functions

Definition 15.1. (*differentiability of vector-valued functions*)
Let $f:[a,b]\to\mathbb{R}^k$ be a vector-valued function and $x\in[a,b]$.
The function f', still vector-valued, is defined by

$$f'(x)=\begin{cases}\lim\limits_{t\to x}\dfrac{f(t)-f(x)}{t-x}, & \text{if } x\in(a,b),\\ \lim\limits_{t\to x^+}\dfrac{f(t)-f(x)}{t-x}, & \text{if } x=a,\\ \lim\limits_{t\to x^-}\dfrac{f(t)-f(x)}{t-x}, & \text{if } x=b,\end{cases}\qquad \text{provided the limit exists.}$$

One can define $f'(x)$ to be the vector such that $f(x+h)=f(x)+f'(x)h+o(h)$
for all h near 0 in $\mathbb{R}$, where $o(h)$ is a vector-valued function with $\lim_{h\to 0}\frac{o(h)}{h}=0$.
The domain of f' is the set of points x at which the limit exists.
The function f' is called the ***derivative of*** *f.*
If f' is defined at x, we say that f is ***differentiable at*** *x.*
If f is differentiable at every point of a set $E\subset[a,b]$,
we say that f is ***differentiable on*** *E.*
Higher-order derivatives are defined in a similar way.

Remark 15.2. (*differentiability of components*)
Let $f=(f_1,\cdots,f_k)$. Then f is differentiable at x
$\Longleftrightarrow$ each f_i is differentiable at x for $1\le i\le k$; in this case $f'(x)=(f_1'(x),\cdots,f_k'(x))$.

Proposition 15.3. (*derivative of scalar product, inner product and cross product*)
Let $f,g:\mathbb{R}\to\mathbb{R}^k$ and $\alpha:\mathbb{R}\to\mathbb{R}$ be differentiable. Then the followings hold (exercise).

1. $(\alpha f)'=\alpha' f+\alpha f'$;
2. $(f\cdot g)'=f'\cdot g+f\cdot g'$;
3. *if $k=3$, then $(f\times g)'=f'\times g+f\times g'$.*

Remark 15.4. (*tangent line and others*)
Let $f:\mathbb{R}\to\mathbb{R}^k$ be differentiable.

1. *If $f(t)$ is the position of a particle moving in $\mathbb{R}^k$ at time t,*
   then $f'(t)$ represents the ***velocity*** *of the particle at time t.*
2. *If $f(t)=a+t(b-a)$, where $a,b\in\mathbb{R}^k$,*
   then f gives a parametric representation of the ***straight line***
   through the points a and b in $\mathbb{R}^k$.
3. *If f gives a parametric representation of a curve in $\mathbb{R}^k$ and $f'(a)\ne 0$,*
   where $a\in\mathbb{R}$ and 0 is the zero vector,
   then the function $g(t)=f(a)+tf'(a)$ gives a parametric representation of
   the ***tangent line*** *to the curve at the point $f(a)$.*
   If $f'(a)=0$, then the curve may not have a tangent line at $f(a)$.

Proposition 15.5.
Let $f:\mathbb{R}\to\mathbb{R}^k$ be differentiable.

1. *There are vector-valued functions for which the ordinary mean value theorem dose not hold.*
2. *If there exists $M>0$ such that $\|f'(x)\|\le M$ for all $x\in[a,b]$,*
   then $\|f(b)-f(a)\|\le M|b-a|$.
3. *Later, we will have* ***variant versions of the mean value theorem***
   in Theorems 19.1 (for $f:\mathbb{R}^k\to\mathbb{R}$) and Theorem 27.1 (for $f:\mathbb{R}^k\to\mathbb{R}^m$).

16 Differentiability in several variables and gradient

Definition 16.1. (*partial derivatives*)
Let $D \subset \mathbb{R}^k$ be open and $f : D \to \mathbb{R}$ be real-valued.
*For any $1 \le i \le k$, the i-**th partial derivative of** f **at a point** $a = (a_1, a_2, \ldots, a_k) \in D$ is defined to be the limit*

$$\partial_i f(a) = \lim_{h \to 0, h \in \mathbb{R}} \frac{f(a_1, \ldots, a_{i-1}, a_i + h, a_{i+1}, \ldots, a_k) - f(a_1, \ldots, a_k)}{h} \quad \text{provided the limit exists.}$$

Other notations for partial derivatives are

$$\partial_i f(a) = \frac{\partial f(a)}{\partial x_i} = \frac{\partial f}{\partial x_i}(a) = \frac{\partial}{\partial x_i} f(a) = \partial_{x_i} f(a) = f_{x_i}(a) = f_i(a).$$

*We say that f is **differentiable** at $a \in D$*
if there exists a vector $m \in \mathbb{R}^k$ such that $\displaystyle\lim_{h \to 0, h \in \mathbb{R}^k} \frac{f(a+h) - f(a) - m \cdot h}{\|h\|} = 0$, where $\cdot$ denotes the inner product.
In this case, the vector m is uniquely determined, and
*we call it the **gradient** of f at a and denote it by $\triangledown f(a)$.*
One can also define $\triangledown f(a)$ to be the vector
such that $f(a+h) = f(a) + \triangledown f(a) \cdot h + o(h)$ for all vectors h close to 0 in $\mathbb{R}^k$,
*where $o(h)$ is a real-valued function of h with $\displaystyle\lim_{h \to 0, h \in \mathbb{R}^k} \frac{o(h)}{\|h\|} = 0$ and is said to be **little** o.*

Remark 16.2. (*gradient, tangent plane, and tangent hyperplane*)
Let $f : \mathbb{R}^k \to \mathbb{R}$ be differentiable at $a \in \mathbb{R}^k$. In Corollary 17.5 and Theorem 18.4, we will have the followings.

1. *The negative gradient $- \triangledown f(a)$ gives a vector in the domain*
   pointing in the direction of the water flow (the steepest descent) at the point,
   which is also orthogonal to the level curve at the point,
   and its magnitude $\| \triangledown f(a)\|$ gives the speed (the steepness) of the water flow at that point;
   while the directional derivative gives the steepness in a given direction.

2. *If $k = 2$, then $z = f(x)$ represents a **surface** in $\mathbb{R}^3$*
   *and $z = f(a) + \triangledown f(a) \cdot (x - a)$ represents the **tangent plane** to the surface $z = f(x)$ at $x = a$,*
   that is, the plane $\{(x, z) \in \mathbb{R}^2 \times \mathbb{R} : (- \triangledown f(a), 1) \cdot ((x, z) - (a, f(a)) = 0\}$ in $\mathbb{R}^3$
   through the point $(a, f(a))$ with the normal vector $(- \triangledown f(a), 1)$.

3. *If $k > 2$, then $z = f(x)$ represents a **hypersurface** in $\mathbb{R}^{k+1}$*
   *and $z = f(a) + \triangledown f(a) \cdot (x - a)$ represents the **tangent hyperplane***
   to the hypersurface $z = f(x)$ at $x = a$.

Theorem 16.3. (*differentiability, continuity, and partial derivatives*)
Let $f : \mathbb{R}^k \to \mathbb{R}$ be a function and $a \in \mathbb{R}^k$.
If f is differentiable at a,
then f is continuous at a, the partial derivatives $\partial_i f(a)$ all exists for $1 \le i \le k$,
and $\triangledown f(a) = (\partial_1 f(a), \cdots, \partial_k f(a))$.

Theorem 16.4. ($C^1 \Rightarrow$ *differentiable* $\Rightarrow$ *partial derivatives exist*)
Let $S \subset \mathbb{R}^k$ be open, $a \in S$, and $f : S \to \mathbb{R}$.

1. *If the partial derivatives $\partial_i f(x)$ all exist for all x in some neighborhood of a*
   and all are continuous at a
   *(in this case we say that f **is** C^1 **at** a or $f \in C^1(\{a\})$),*
   then f is differentiable at a and $\triangledown f(a) = (\partial_1 f(a), \cdots, \partial_k f(a))$.

2. *If the partial derivatives $\partial_i f(x)$ all exist and are bounded for all $x \in S$,*
   then f is continuous on S;
   however, it is possible that f is not differentiable at a point in S when $k \ge 2$. (exercise)

17 Differential and directional derivative

Definition 17.1.
Let $f : \mathbb{R}^k \to \mathbb{R}$ be differentiable and $a \in \mathbb{R}^k$.
Let $h = (h_1, h_2, \cdots, h_k)$ be the given increment from a to $a + h$.
*Then the **increment** of f at a is given by $\triangle f(a; h) := f(a + h) - f(a) = \triangledown f(a) \cdot h + o(h)$,*
which represents the changes in the value of f when a changes to $a + h$.
If the increment $h = (dx_1, dx_2, \cdots, dx_k)$ is infinitesimal,
then the increment of f at a can be linearly approximated by $\triangledown f(a) \cdot h$ by neglecting the error term $o(h)$.
*This leads us to define the **differential** of f at x as $df(x) := \triangledown f(x) \cdot (dx_1, dx_2, \cdots, dx_k) = \partial_1 f(x)dx_1 + \cdots + \partial_k f(x)dx_k$.*
As applications, the following questions can be answered about a circular cylinder with base radius r and height h.
(i) About (i.e., approximately) how much does the volume increases if (r, h) is increased from $(1, 2)$ to $(1.02, 2.01)$?
(ii) About how much should r be decreased from 1 to keep the volume constant, if h is increased from 2 to 2.01?

Remark 17.2. (*sum, product, quotient rules of differentials*)
Let $f, g : \mathbb{R}^k \to \mathbb{R}$ be differentiable.
Then $d(f \pm g) = df \pm dg$, $d(fg) = fdg + gdf$,
and $d\left(\frac{f}{g}\right) = \frac{gdf - fdg}{g^2}$ at which g does not vanish.

Definition 17.3. (*directional derivative*)
Let $f : \mathbb{R}^k \to \mathbb{R}$ be a function and $a \in \mathbb{R}^k$ be a point.
*Let u be a **unit vector** in $\mathbb{R}^k$, that is, $\|u\| = 1$.*
*The **directional derivative of** f **at** a **in the direction of** u is defined to be*

$$\partial_u f(a) = \frac{d}{dt} f(a + tu)\Big|_{t=0} = \lim_{t \to 0, t \in \mathbb{R}} \frac{f(a + tu) - f(a)}{t} \text{ provided the limit exists.}$$

Theorem 17.4. (*differentiability and directional derivative*)
Let $f : \mathbb{R}^k \to \mathbb{R}$ be differentiable at $a \in \mathbb{R}^k$.
Then the directional derivative $\partial_u f(a)$ exists for all unit vectors $u \in \mathbb{R}^k$,
and $\partial_u f(a) = \triangledown f(a) \cdot u$.
But the existence of all directional derivatives dose not imply differentiability. (exercise)

Corollary 17.5. (*gradient and extreme values of directional derivatives*)
Let $f : \mathbb{R}^k \to \mathbb{R}$ be differentiable at the point $a \in \mathbb{R}^k$ where $\triangledown f(a) \neq 0$,
then the maximal (resp. minimal) value of $\partial_u f(a)$ over all unit vectors u
is $\| \triangledown f(a)\|$ (resp. $-\| \triangledown f(a)\|$) and it occurs when $u = \frac{\triangledown f(a)}{\|\triangledown f(a)\|}$ (resp. $u = -\frac{\triangledown f(a)}{\|\triangledown f(a)\|}$).

18 The chain rule

Theorem 18.1 (the chain rule for one variable).
Let $g : \mathbb{R} \to \mathbb{R}^k$ *be a differentiable at* $a \in \mathbb{R}$ *and*
let $f : \mathbb{R}^k \to \mathbb{R}$ *be a differentiable at* $g(a) \in \mathbb{R}^k$.
Then the composition function $f \circ g$ *is differentiable at* a *and* $(f \circ g)'(a) = \triangledown f(g(a)) \cdot g'(a)$.
In Leibniz's notation, with $t \mapsto z(x_1(t), x_2(t), \cdots, x_k(t))$,
we have $\dfrac{dz}{dt}(a) = \displaystyle\sum_{i=1}^{k} \frac{\partial z}{\partial x_i}(x_1(a), x_2(a), \cdots, x_k(a)) \frac{dx_i}{dt}(a)$.
Writing explicit variables is to avoid confusion of z' *and* $\partial_i z$ *on both sides.*

Theorem 18.2 (the chain rule for several variables).
Let $g = (g_1, \cdots, g_k) : \mathbb{R}^m \to \mathbb{R}^k$ *be a function such that*
for each $1 \le i \le k$, *the map* $g_i : \mathbb{R}^m \to \mathbb{R}$ *is differentiable at* $a \in \mathbb{R}^m$ *(resp. of* C^1 *near* a*),*
and let $f : \mathbb{R}^k \to \mathbb{R}$ *be a differentiable at* $g(a) \in \mathbb{R}^k$ *(resp. of* C^1 *near* $g(a)$*).*
Then the composition function $f \circ g$ *is differentiable at* a *(resp. of* C^1 *near* a*)*
and $\partial_j (f \circ g)(a) = \triangledown f(g(a)) \cdot (\partial_j g_1(a), \cdots, \partial_j g_k(a))$ *for all* $1 \le j \le m$.
In Leibniz's notation, with $(t_1, \cdots, t_m) \mapsto z(x_1(t_1, \cdots, t_m), \cdots, x_k(t_1, \cdots, t_m))$,
we have $\dfrac{\partial z}{\partial t_j}(a) = \displaystyle\sum_{i=1}^{k} \frac{\partial z}{\partial x_i}(x_1(a), \cdots, x_k(a)) \frac{\partial x_i}{\partial t_j}(a)$ *for all* $j = 1, \ldots, m$.
Writing explicit variables is to avoid confusion of $\partial_j z$ *and* $\partial_i z$ *on both sides.*

Theorem 18.3 (Euler's theorem).
Let $f : \mathbb{R}^k \to \mathbb{R}$ *be a function which is* ***positively homogeneous of degree*** $r \in \mathbb{R}$,
that is, $f(tx) = t^r f(x)$ *for all* $t > 0$ *and* $x \neq 0$, *e.g.,* $f(x_1, x_2, x_3) = x_1^3 x_2^2 x_3^1 + x_1^6$.
If, in addition, f *is differentiable at* $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$ *with* $a \neq 0$,
then $rf(a) = a \cdot \triangledown f(a) = \sum_{i=1}^{k} a_i \partial_i f(a)$.

Theorem 18.4. (*gradient, normal, and orthogonality*)

1. *Let* $U \subset \mathbb{R}^3$ *be open and* $f : U \to \mathbb{R}$ *be differentiable on* U *such that*
   $S = \{x \in U : f(x) = 0\}$ *is a two-dimensional smooth surface in* $\mathbb{R}^3$.
   If $a \in S$ *and* $\triangledown f(a) \neq 0$, *then the vector* $\triangledown f(a)$ *is normal to the surface* S *at* a,
   and the equation of the tangent plane to S *at* a *is* $\triangledown f(a) \cdot (x - a) = 0$.

2. *Let* $V \subset \mathbb{R}^2$ *be open,* $g : V \to \mathbb{R}$ *be differentiable on* V,
   $T = \{(x, z) \in \mathbb{R}^2 \times \mathbb{R} : x \in V \text{ and } z = g(x)\}$, *and* $a \in V$.
   Then the tangent plane to T *at* $(a, g(a))$ *is given by the equation* $z = g(a) + \triangledown g(a) \cdot (x - a)$,
   that is, the plane in $\mathbb{R}^2 \times \mathbb{R}$ *through the point* $(a, g(a))$ *with the normal vector* $(-\triangledown g(a), 1)$.

3. *Let* $V \subset \mathbb{R}^2$ *be open,* $g : V \to \mathbb{R}$ *be differentiable on* V, *and* $(a_1, a_2) \in V$.
   if a differentiable function $t \mapsto (x_1(t), x_2(t)) \in V$ *for* $-1 < t < 1$
   parameterizes ***a part of the level curve*** *of* g *at* (a_1, a_2),
   that is, $\{(x_1(t), x_2(t)) \in V : -1 < t < 1\} \subset \{(x_1, x_2) \in V : g(x_1, x_2) = g(a_1, a_2)\}$ *and* $(x_1(0), x_2(0)) = (a_1, a_2)$,
   then $\triangledown g(a_1, a_2) \cdot (x_1'(0), x_2'(0)) = 0$,
   that is, the gradient is orthogonal to the level curve at the given point.

Proposition 18.5.
Let $F, G : \mathbb{R}^m \to \mathbb{R}$ *be differentiable.*
Derive the formulas for the partial derivatives of $F + G$, FG, F/G, *and* F^G, *by using Theorem 18.2.* (*exercise*)

19 The mean value theorem and Rolle's theorem in several variables

Theorem 19.1 (the mean value theorem in several variables)**.**
Let $a, b \in \mathbb{R}^k$ with $a \neq b$ and let $L = \{a + t(b-a) : 0 \le t \le 1\}$, which is the line segment from a to b.
Let $S \subset \mathbb{R}^k$ be an open set with $L \subset S$.
If $f : S \to \mathbb{R}$ is continuous on L and is differentiable on $L \backslash \{a, b\}$,
then there exists $c \in L$ such that $f(b) - f(a) = \triangledown f(c) \cdot (b-a)$, where a, b are written as row vectors in $\mathbb{R}^k$.

Corollary 19.2.
*Let $S \subset \mathbb{R}^k$ be open and **convex**, that is, $\{a + t(b-a) : 0 \le t \le 1\} \subset S$ for any $a, b \in S$.*
Let $f : S \to \mathbb{R}$ be differentiable on S.

1. *If there exists $M > 0$ such that $\| \triangledown f(x)\| \le M$ for every $x \in S$,*
   then $|f(b) - f(a)| \le M\|b - a\|$ for all $a, b \in S$.
   The convexity is necessary here.

2. *If $\triangledown f(x) = 0$ for all $x \in S$, then f is a constant function on S.*
   This result is generalized in the following theorem.

Theorem 19.3.
Let $S \subset \mathbb{R}^k$ be open and connected.
If $f : S \to \mathbb{R}$ is differentiable on S and $\triangledown f(x) = 0$ for all $x \in S$,
then f is a constant function on S.
The connectedness is necessary here.

Theorem 19.4 (Rolle's theorem in several variables)**.**

1. *Let $a, b \in \mathbb{R}^k$ with $a \neq b$, $L = \{a + t(b-a) : 0 \le t \le 1\}$, and $S \subset \mathbb{R}^k$ be an open set with $L \subset S$.*
   If $f : S \to \mathbb{R}$ is continuous on L and is differentiable on $L \backslash \{a, b\}$ with $f(a) = f(b)$,
   then there exists $c \in L \backslash \{a, b\}$ such that $\partial_u f(c) = 0$, where $u = \frac{b-a}{\|b-a\|}$. (*exercise*)

2. *Let $S \subset \mathbb{R}^k$ be nonempty, open and bounded.*
   Let $f : \mathbb{R}^k \to \mathbb{R}$ be continuous on $\overline{S}$, be differentiable on S, and be constant on ∂S.
   Then there exists $c \in S$ such that $\triangledown f(c) = 0$. (*exercise*)

Proposition 19.5.
Let $S \subset \mathbb{R}^k$ be open and connected. Let $f : S \to \mathbb{R}$ be differentiable on S. (*exercise*)

1. *If $k = 1$ and $f'(x) = 0$ for all $x \in S$, then f is a constant function on S.*

2. *If $k \ge 2$, S is convex and $\partial_1 f(x) = 0$ for all $x \in S$,*
   *then f is **a constant function in the first variable**, that is, if $(x_1, a), (x_2, a) \in S$ then $f(x_1, a) = f(x_2, a)$.*

3. *The convexity is necessary in Item 2.*

20 Implicit differentiation

Theorem 20.1. (*implicit differentiation*)

1. *If a differentiable function* $y = y(x_1, \cdots, x_k)$ *is defined implicitly by the equation* $F(x_1, \cdots, x_k, y) = 0$, *where* F *is differentiable, then for every* $1 \le j \le k$,

$$\frac{\partial y}{\partial x_j}(x_1, \cdots, x_k) = -\frac{\frac{\partial F}{\partial x_j}(x_1, \cdots, x_k, y)}{\frac{\partial F}{\partial y}(x_1, \cdots, x_k, y)} \quad \textit{provided} \quad \frac{\partial F}{\partial y}(x_1, \cdots, x_k, y) \neq 0.$$

2. *If two differentiable functions* $u = u(x, y)$ *and* $v = v(x, y)$ *are defined implicitly by the equations* $F(x, y, u, v) = 0$ *and* $G(x, y, u, v) = 0$, *where* F *and* G *are differentiable, then*

$$\frac{\partial u}{\partial x} = -\frac{det\begin{bmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial v} \\ \frac{\partial G}{\partial x} & \frac{\partial G}{\partial v} \end{bmatrix}}{det\begin{bmatrix} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial v} \\ \frac{\partial G}{\partial u} & \frac{\partial G}{\partial v} \end{bmatrix}} \quad \textit{and} \quad \frac{\partial v}{\partial x} = -\frac{det\begin{bmatrix} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial x} \\ \frac{\partial G}{\partial u} & \frac{\partial G}{\partial x} \end{bmatrix}}{det\begin{bmatrix} \frac{\partial F}{\partial u} & \frac{\partial F}{\partial v} \\ \frac{\partial G}{\partial u} & \frac{\partial G}{\partial v} \end{bmatrix}},$$

provided the denominator is not zero.

3. *If there are* k *differentiable functions* $y_i = y_i(x_1, \cdots, x_m)$, $1 \le i \le k$ *defined implicitly by the equations* $F_i(x_1, \cdots, x_m, y_1, \cdots, y_k) = 0$, *where* F_i *is differentiable for all* $1 \le i \le k$, *then for* $1 \le i \le k$ *and* $1 \le j \le m$, *we have*

$$\frac{\partial y_i}{\partial x_j} = -\frac{det\begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial x_j} & \cdots & \frac{\partial F_1}{\partial y_k} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial F_i}{\partial y_1} & \cdots & \frac{\partial F_i}{\partial x_j} & \cdots & \frac{\partial F_i}{\partial y_k} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial F_k}{\partial y_1} & \cdots & \frac{\partial F_k}{\partial x_j} & \cdots & \frac{\partial F_k}{\partial y_k} \end{bmatrix}}{det\begin{bmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \cdots & \cdots & \frac{\partial F_1}{\partial y_k} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial F_i}{\partial y_1} & \cdots & \cdots & \cdots & \frac{\partial F_i}{\partial y_k} \\ \vdots & & \vdots & & \vdots \\ \frac{\partial F_k}{\partial y_1} & \cdots & \cdots & \cdots & \frac{\partial F_k}{\partial y_k} \end{bmatrix}},$$

provided the denominator is not zero.

Remark 20.2.

1. *When practicing implicit differentiations, we should go through the proof of the above theorems, instead of directly applying the formulas.*

2. *If* y *is implicitly defined by a function* $F(x_1, \ldots, x_k, y) = 0$ *and the variables* x_i*'s are constrained to satisfy a equation* $H(x_1, \ldots, x_k) = 0$, *then the differentiation of* y *with respect to* x_1, *depends critically on which variable is fixed and the other* $k-1$ *variables, including* x_1, *are considered as independent variables, e.g.,* $\frac{\partial y}{\partial x_1}|_{\{x_k \text{ is fixed}\}}$ *and* $y(x_1, x_2 \ldots, x_{k-1})$ *or* $\frac{\partial y}{\partial x_1}|_{\{x_2 \text{ is fixed}\}}$ *and* $y(x_1, x_3 \ldots, x_k)$.

21 Higher-order partial derivatives, the Laplacian, and the multinomial theorem

Definition 21.1. (C^r *and* C^∞ *functions*)
Let $U \subset \mathbb{R}^k$ *be open,* $f : U \to \mathbb{R}$*, and* $r \in \mathbb{N}$.
The standard notations for the second derivative are

$$\frac{\partial}{\partial x_j}\left[\frac{\partial f}{\partial x_i}\right] = \frac{\partial^2 f}{\partial x_j \partial x_i} = \partial_{x_j}\partial_{x_i} f = \partial^2_{ji} f = \partial_j \partial_i f = \partial_{x_i x_j} f = f_{ij}, \text{ and } \frac{\partial}{\partial x_i}\left[\frac{\partial f}{\partial x_i}\right] = \frac{\partial^2 f}{\partial x_i^2} = \partial_i^2 f.$$

We say that f *is of* ***class*** C^r *on* U
if all of its partial derivatives of order $\le r$ *exist and are continuous on* U.
We say that f *is of* ***class*** C^∞ *on* U
if all of its partial derivatives of all orders exist and are continuous on U.

Theorem 21.2. (*reordering of partial derivatives*)
Let $S \subset \mathbb{R}^k$ *be open,* $f : S \to \mathbb{R}$*,* $a \in S$*, and* $1 \le i, j \le k$.
If $\partial_i f, \partial_j f$*, and* $\partial^2_{ji} f$ *exist on* S*, and* $\partial^2_{ji} f$ *is continuous at* a,
then $\partial^2_{ij} f(a)$ *exists and* $\partial^2_{ij} f(a) = \partial^2_{ji} f(a)$.
For $\partial^2_{ij} f(a) \ne \partial^2_{ji} f(a)$*, refer to the example* $f(x,y) = \begin{cases} 0, & \text{if } (x,y) = (0,0), \\ \frac{xy(x^2-y^2)}{x^2+y^2}, & \text{otherwise.} \end{cases}$

Corollary 21.3. (C^r *and reordering of partial derivatives*)
Let $S \subset \mathbb{R}^k$ *be open and* $f : S \to \mathbb{R}$.

1. *If* f *is* C^2 *on* S*, then* $\partial^2_{ji} f = \partial^2_{ij} f$ *on* S *for all* $1 \le i, j \le k$.
2. *If* f *is* C^r *on* S *for some* $r \in \mathbb{N}$*, then* $\partial^r_{i_1 i_2 \cdots i_r} f = \partial^r_{j_1 j_2 \cdots j_r} f$ *on* S *for all* $1 \le i_1, i_2, \cdots, i_r \le k$,
   where $(j_1, j_2, \cdots, j_r)$ *is a reordering of* $(i_1, i_2, \cdots, i_r)$.

Proposition 21.4. (*the Laplacian of* f)
Let $S \subset \mathbb{R}^2$ *be open and let* $f : S \to \mathbb{R}$ *be* C^2 *on* S *and be denoted by* $(x,y) \mapsto f(x,y)$.
The ***Laplacian*** *of* f *is defined to be* $\dfrac{\partial^2 f}{\partial x^2} + \dfrac{\partial^2 f}{\partial y^2}$.
If $x = r\cos\theta$ *and* $y = r\sin\theta$,
then $\dfrac{\partial^2 f}{\partial x^2} + \dfrac{\partial^2 f}{\partial y^2} = \dfrac{\partial^2 f}{\partial r^2} + \dfrac{1}{r}\dfrac{\partial f}{\partial r} + \dfrac{1}{r^2}\dfrac{\partial^2 f}{\partial \theta^2}$

Definition 21.5.
For a k*-tuple* $\alpha = (\alpha_1, \cdots, \alpha_k)$ *of nonnegative integers, we call it a* ***multi-index***
and define the sum of components $|\alpha|_c = \alpha_1 + \alpha_2 + \cdots + \alpha_k$*, the product of components* $\alpha! = \alpha_1! \alpha_2! \cdots \alpha_k!$,
$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$*, where* $x = (x_1, x_2, \cdots, x_k) \in \mathbb{R}^k$,
and $\partial^\alpha f = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_k^{\alpha_k} f = \frac{\partial^{|\alpha|_c} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_k^{\alpha_k}}$*, where both* ∂_i^0 *and* ∂x_i^0 *are conventionally neglected.*
Notice that $\alpha \in \mathbb{N}_0^k \subset \mathbb{R}^k$ *and* $|\alpha|_c \ne \|\alpha\|$ *when* $k \ge 2$.

Theorem 21.6 (the multinomial theorem)**.**
For any $x = (x_1, x_2, \cdots, x_k) \in \mathbb{R}^k$ *and* $n \in \mathbb{N}$,
we have $(x_1 + x_2 + \cdots + x_k)^n = \displaystyle\sum_{\alpha \in \mathbb{N}_0^k, |\alpha|_c = n} \frac{n!}{\alpha!} x^\alpha = \sum_{\substack{\alpha_1 + \alpha_2 + \cdots + \alpha_k = n \\ 0 \le \alpha_1, \alpha_2, \cdots, \alpha_k \le n}} \frac{n!}{\alpha_1! \alpha_2! \cdots \alpha_k!} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$.
If $k = 2$*, the above result is called* the binomial theorem.

Theorem 21.7 (the high-dimensional binomial theorem)**.** (*exercise*)
For two vectors $x = (x_1, \cdots, x_k), y = (y_1, \cdots, y_k) \in \mathbb{R}^k$ *and a multi-index* $\gamma = (\gamma_1, \cdots, \gamma_k) \in \mathbb{N}_0^k$,
we have $(x+y)^\gamma = \displaystyle\sum_{\alpha, \beta \in \mathbb{N}_0^k, \alpha + \beta = \gamma} \frac{\gamma!}{\alpha! \beta!} x^\alpha y^\beta = \sum_{\substack{\alpha_i + \beta_i = \gamma_i \text{ for all } 1 \le i \le k \\ 0 \le \alpha_i, \beta_i \le \gamma_i}} \frac{\gamma_1! \cdots \gamma_k!}{\alpha_1! \cdots \alpha_k! \beta_1! \cdots \beta_k!} x_1^{\alpha_1} \cdots x_k^{\alpha_k} y_1^{\beta_1} \cdots y_k^{\beta_k}$.

Proposition 21.8 (the product rule for partial derivatives)**.** (*exercise*)
Let $S \subset \mathbb{R}^k$ *be open,* $f, g : S \to \mathbb{R}$ *be* C^r *differentiable on* S*, and* $\gamma = (\gamma_1, \cdots, \gamma_k) \in \mathbb{N}_0^k$.
Then $\partial^\gamma (fg) = \displaystyle\sum_{\alpha, \beta \in \mathbb{N}_0^k, \alpha + \beta = \gamma} \frac{\gamma!}{\alpha! \beta!} (\partial^\alpha f)(\partial^\beta g) = \sum_{\substack{\alpha_i + \beta_i = \gamma_i \text{ for all } 1 \le i \le k \\ 0 \le \alpha_i, \beta_i \le \gamma_i}} \frac{\gamma_1! \cdots \gamma_k!}{\alpha_1! \cdots \alpha_k! \beta_1! \cdots \beta_k!} (\partial_1^{\alpha_1} \partial_2^{\alpha_2} \cdots \partial_k^{\alpha_k} f)(\partial_1^{\beta_1} \partial_2^{\beta_2} \cdots \partial_k^{\beta_k} g)$.

22 Taylor's theorem in one variable

Theorem 22.1 (Taylor's theorem I with integral remainder)**.**
Let $f : I \to \mathbb{R}$ be C^{m+1} on I and $c \in I$,
where $I \subset \mathbb{R}$ is an open interval and $m \geq 0$ is an integer.
Then for any $x \in I$,

$$f(x) = \sum_{n=0}^{m} \frac{f^{(n)}(c)}{n!}(x-c)^n + \frac{\int_0^1 (1-t)^m f^{(m+1)}\left(c + t(x-c)\right) dt}{m!}(x-c)^{m+1};$$

here conventionally we write $f^{(0)}(c) = f(c)$.
The summation term $\sum_{n=0}^{m} \frac{f^{(n)}(c)}{n!}(x-c)^n$ is called
*the m**th-order Taylor polynomial** for f about c on I.*
The difference $f(x) - \sum_{n=0}^{m} \frac{f^{(n)}(c)}{n!}(x-c)^n$ is called
*the m**th-order Taylor remainder** for f about c on I and is denoted by $R_{f,c,m}(x-c)$.*

If, in addition, there exists $M \in [0,\infty)$ such that $|f^{(m+1)}(y)| \leq M$ for all $y \in I$,
then for any $x \in I$, we have $|R_{f,c,m}(x-c)| \leq \frac{M}{(m+1)!}|x-c|^{m+1}$.

Theorem 22.2 (Taylor's theorem II with integral remainder)**.**
Let $f : I \to \mathbb{R}$ be C^m on I and $c \in I$,
where $I \subset \mathbb{R}$ is an open interval and $m \geq 1$ is an integer.
Then for any $x \in I$,

$$f(x) = \sum_{n=0}^{m} \frac{f^{(n)}(c)}{n!}(x-c)^n + \frac{\int_0^1 (1-t)^{m-1} \left[f^{(m)}\left(c + t(x-c)\right) - f^{(m)}(c)\right] dt}{(m-1)!}(x-c)^m$$

and the mth-order Taylor remainder for f about c satisfies $\lim_{x \to c} \frac{R_{f,c,m}(x-c)}{(x-c)^m} = 0$.

Theorem 22.3 (Taylor's theorem with Lagrange's remainder and Taylor series)**.**
Let $f : I \to \mathbb{R}$ be differentiable of order $m+1$ on I and $c \in I$,
where $I \subset \mathbb{R}$ is an open interval and $m \geq 1$ is an integer.
Then for any $x \in I$, there exists z between x and c such that

$$f(x) = \sum_{n=0}^{m} \frac{f^{(n)}(c)}{n!}(x-c)^n + \frac{f^{(m+1)}(z)}{(m+1)!}(x-c)^{m+1}.$$

If, in addition, f has derivatives of all orders on I
and $\lim_{m \to \infty} \frac{f^{(m+1)}(z)}{(m+1)!}(x-c)^{m+1} = 0$ for all $x \in I$ and all z between x and c,
then for any $x \in I$, $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n$
*which is a convergent series and is called the **Taylor series** for f about c on I.*
Notice that if there exists $M \in [0,\infty)$ such that $|f^{(n)}(y)| \leq M^n$ for all $y \in I$ and integers $n \geq 0$,
then $\lim_{m \to \infty} \frac{f^{(m+1)}(z)}{(m+1)!}(x-c)^{m+1} = 0$ for all $x \in I$ and all z between x and c,,

Example 22.4. (*the Taylor series for elementary functions*)
$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$ *on* $(-1,1)$, $e^x = \sum_{n=0}^{\infty} \frac{1}{n!}x^n$ *on* $\mathbb{R}$, $\ln(1+x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)} x^{n+1}$ *on* $(-1,1]$,
$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}$ *on* $\mathbb{R}$, *and* $\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}$ *on* $\mathbb{R}$.

23 Taylor's theorem in several variables

Theorem 23.1 (Taylor's theorem in several variables).
Let $S \subset \mathbb{R}^k$ be an open convex set, $c = (c_1, c_2, \cdots, c_k) \in S$, and $m \geq 1$ be an integer.

1. *If $f : S \to \mathbb{R}$ be C^m on S,*
   then for any $x = (x_1, x_2, \cdots, x_k) \in S$,

$$f(x) = \sum_{\substack{n=0 \\ \alpha_1+\alpha_2+\cdots+\alpha_k=n \\ 0\leq\alpha_1,\alpha_2,\cdots,\alpha_k\leq n}}^{m} \frac{\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(c)}{\alpha_1!\alpha_2!\cdots\alpha_k!}(x_1-c_1)^{\alpha_1}(x_2-c_2)^{\alpha_2}\cdots(x_k-c_k)^{\alpha_k} + R_{f,c,m}(x-c),$$

   where

$$\begin{aligned} &R_{f,c,m}(x-c) \\ &= \sum_{\substack{\alpha_1+\alpha_2+\cdots+\alpha_k=m \\ 0\leq\alpha_1,\alpha_2,\cdots,\alpha_k\leq m}} \frac{\int_0^1 (1-t)^{m-1}\left[\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(c+t(x-c)) - \partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(c)\right]dt}{\frac{\alpha_1!\alpha_2!\cdots\alpha_k!}{m}}(x_1-c_1)^{\alpha_1}(x_2-c_2)^{\alpha_2}\cdots(x_k-c_k)^{\alpha_k} \end{aligned}$$

   and satisfies $\lim_{x\to c}\frac{R_{f,c,m}(x-c)}{\|x-c\|^m} = 0$.

   If, in addition, there exits $M_1 > 0$ and $\lambda > 0$ such that for any $0 \leq \alpha_1, \alpha_2, \cdots, \alpha_k \leq m$ satisfying $\alpha_1 + \alpha_2 + \cdots + \alpha_k = m$ and for any $x, y \in S$,

$$|\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(x) - \partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(y)| \leq M_1\|x-y\|^{\lambda},$$

   then there exists $M_2 > 0$ such that $|R_{f,c,m}(x-c)| \leq M_2\|x-c\|^{\lambda+m}$ for all $x \in S$.

2. *If $f : S \to \mathbb{R}$ be C^{m+1} on S with $m \geq 1$,*
   then for any $x = (x_1, x_2, \cdots, x_k) \in S$,

$$f(x) = \sum_{\substack{n=0 \\ \alpha_1+\alpha_2+\cdots+\alpha_k=n \\ 0\leq\alpha_1,\alpha_2,\cdots,\alpha_k\leq n}}^{m} \frac{\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(c)}{\alpha_1!\alpha_2!\cdots\alpha_k!}(x_1-c_1)^{\alpha_1}(x_2-c_2)^{\alpha_2}\cdots(x_k-c_k)^{\alpha_k} + R_{f,c,m}(x-c),$$

   where

$$R_{f,c,m}(x-c) = \sum_{\substack{\alpha_1+\alpha_2+\cdots+\alpha_k=m+1 \\ 0\leq\alpha_1,\alpha_2,\cdots,\alpha_k\leq m+1}} \frac{\int_0^1 (1-t)^{m}\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(c+t(x-c))dt}{\frac{\alpha_1!\alpha_2!\cdots\alpha_k!}{m+1}}(x_1-c_1)^{\alpha_1}(x_2-c_2)^{\alpha_2}\cdots(x_k-c_k)^{\alpha_k}.$$

   and we also have that there exists z between x and c such that

$$R_{f,c,m}(x-c) = \sum_{\substack{\alpha_1+\alpha_2+\cdots+\alpha_k=m+1 \\ 0\leq\alpha_1,\alpha_2,\cdots,\alpha_k\leq m+1}} \frac{\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(z)}{\alpha_1!\alpha_2!\cdots\alpha_k!}(x_1-c_1)^{\alpha_1}(x_2-c_2)^{\alpha_2}\cdots(x_k-c_k)^{\alpha_k}.$$

   If, in addition, there exists $M \in [0, \infty)$ such that $|\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(y)| \leq M$
   for all $y \in S$ and all $0 \leq \alpha_1, \alpha_2, \cdots, \alpha_k \leq m+1$ with $\alpha_1 + \alpha_2 + \cdots + \alpha_k = m+1$,
   then for any $x \in S$, $|R_{f,c,m}(x-c)| \leq \frac{M}{(m+1)!}(|x_1-c_1| + \cdots + |x_k-c_k|)^{m+1}$.

24 The uniqueness of the Taylor polynomial and the second derivative test in two variables

Theorem 24.1 (the uniqueness of the Taylor polynomial).
Let $S \subset \mathbb{R}^k$ be an open convex set, $c = (c_1, c_2, \cdots, c_k) \in S$, and $m \geq 1$ be an integer.
Let $f : S \to \mathbb{R}$ be C^m on S and $f(x) = Q(x-c) + E(x-c)$ for all $x \in S$,
where $Q : \mathbb{R}^k \to \mathbb{R}$ is a polynomial in $x - c$ of degree $\leq m$
and $E : \mathbb{R}^k \to \mathbb{R}$ is a function with $\lim_{x \to c} \frac{E(x-c)}{\|x-c\|^m} = 0$.
Then for any $x = (x_1, x_2, \cdots, x_k) \in S$,

$$Q(x-c) = \sum_{\substack{n=0 \\ \alpha_1+\alpha_2+\cdots+\alpha_k=n \\ 0\leq\alpha_1,\alpha_2,\cdots,\alpha_k\leq n}}^{m} \frac{\partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_k^{\alpha_k} f(c)}{\alpha_1!\alpha_2!\cdots\alpha_k!}(x_1-c_1)^{\alpha_1}(x_2-c_2)^{\alpha_2}\cdots(x_k-c_k)^{\alpha_k},$$

i.e., $Q(x-c)$ is the mth-order Taylor polynomial for f about c on S.

If $k = 1$, we may only require f is C^{m-1} on S and $f^{(m)}(c)$ exists. (exercise)

Theorem 24.2. (*the critical point theorem in several variables*)
Let $S \subset \mathbb{R}^k$ be open and $f : S \to \mathbb{R}$ be differentiable at $c \in S$.
If f has a local maximum or a local minimum at c,
*then c is a **critical point** of f, i.e., $\nabla f(c) = 0$.*

Theorem 24.3 (the second derivative test in several variables).
Let $S \subset \mathbb{R}^k$ be open, $f : S \to \mathbb{R}$ be C^2 at $c \in S$, and c is a critical point of f.

Let $H_f(c) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(c) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(c) & \cdots & \frac{\partial^2 f}{\partial x_k \partial x_1}(c) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(c) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(c) & \cdots & \frac{\partial^2 f}{\partial x_k \partial x_2}(c) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_k}(c) & \frac{\partial^2 f}{\partial x_2 \partial x_k}(c) & \cdots & \frac{\partial^2 f}{\partial x_k \partial x_k}(c) \end{bmatrix}$, *called the **Hessian** of f at c.*

1. *If all eigenvalues of $H_f(c)$ are positive (resp. negative),*
   then f has a local minimum (resp. maximum) at c.
2. *If f has a local minimum (resp. maximum) at c,*
   then all eigenvalues of $H_f(c)$ are nonnegative (resp. nonpositive).

Theorem 24.4 (the second derivative test in two variables).
Let $S \subset \mathbb{R}^2$ be open, $f : S \to \mathbb{R}$ be C^2 at $c \in S$, and c is a critical point of f.
*Let $H_f(c)$ be the **Hessian** of f at c, that is,* $H_f(c) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(c) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(c) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(c) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(c) \end{bmatrix}$.

1. *If $\det(H_f(c)) > 0$ and $\frac{\partial^2 f}{\partial x_1 \partial x_1}(c) > 0$, then f has a local minimum at c.*
2. *If $\det(H_f(c)) > 0$ and $\frac{\partial^2 f}{\partial x_1 \partial x_1}(c) < 0$, then f has a local maximum at c.*
3. *If $\det(H_f(c)) < 0$,*
   *then c is a **saddle point** of f, that is, there are two eigenvalues of $H_f(c)$ which are of opposite signs.*
4. *If $\det(H_f(c)) = 0$, that is, the critical point c is said to be **degenerate**,*
   then conclusions vary depending high-order derivatives.

Theorem 24.5 (extreme value theorem for C^r functions). (*exercise*)
Let $f : (a,b) \to \mathbb{R}$ be a C^r function on (a,b), where $r \in \mathbb{N}$, and let $c \in (a,b)$.
If $f^{(n)}(c) = 0$ for all $1 \leq n < r$ and $f^{(r)}(c) \neq 0$, then the followings hold.

1. *If r is even and $f^{(r)}(c) > 0$, then f has a local minimum at c.*
2. *If r is even and $f^{(r)}(c) < 0$, then f has a local maximum at c.*
3. *If r is odd, then f has neither a local minimum nor a local maximum at c.*

25 Extreme value problem and Lagrange's method for constraint

Theorem 25.1 (extreme value problem on a closed and unbounded region).
Let $S \subset \mathbb{R}^k$ be closed and unbounded and $f : S \to \mathbb{R}$ be continuous on S.

1. *If $\lim\limits_{\substack{\|x\|\to+\infty \\ x\in S}} f(x) = +\infty$ (resp. $\lim\limits_{\substack{\|x\|\to+\infty \\ x\in S}} f(x) = -\infty$),*
   then f has a global minimum but no global maximum on S
   (resp. a global maximum but no global minimum on S).

2. *If $\lim\limits_{\substack{\|x\|\to+\infty \\ x\in S}} f(x) = 0$ and there is a point $a \in S$ such that $f(a) < 0$ (resp. $f(a) > 0$),*
   then f has a global minimum (resp. a global maximum) on S.

Theorem 25.2 (Lagrange's method for constraints).
Let $S \subset \mathbb{R}^k$ be open and let $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}^m$ be two C^1 functions on S with $m < k$ and write $g = (g_1, \dots, g_m)$.
Let $T = \{x \in S : g(x) = 0\}$ and $a \in T$.
If the restriction of f on T, $f|_T$, has a local maximum or a local minimum at a and the the vectors $\triangledown g_1(a), \dots, \triangledown g_m(a)$ are linearly independent ($\triangledown g_1(a) \neq 0$ for $m = 1$), then there exists m real numbers $\lambda_1, \dots, \lambda_m$ such that

$$\triangledown f(a) + \lambda_1 \triangledown g_1(a) + \cdots + \lambda_m \triangledown g_m(a) = 0.$$

Here the numbers $\lambda_1, \dots, \lambda_m$ are called ***Lagrange's multipliers***.

In practice, one solves the system of $(m + k)$ equations
$g_i(x) = 0$ and $\frac{\partial f(x)}{\partial x_j} + \lambda_1 \frac{\partial g_1(x)}{\partial x_j} + \cdots + \lambda_m \frac{\partial g_m(x)}{\partial x_j} = 0$ for $1 \le i \le m$ and $1 \le j \le k$
for $(m + k)$ variables λ_i and x_j for $1 \le i \le m$ and $1 \le j \le k$, where $x = (x_{1,} \cdots x_k)$,
which are the candidates of local extreme points on T.

26 The Fréchet derivative, Jacobian, and chain rule

Definition 26.1 (Fréchet derivative for vector-valued functions in several variables)**.**
Let $S \subset \mathbb{R}^k$ be open and $f : S \to \mathbb{R}^m$.
*We say that f is **differentiable** at $x \in S$*
if there is a linear transformation A from $\mathbb{R}^k$ to $\mathbb{R}^m$ such that $\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0$;
by linear transformation, we means that $A(c_1 x + c_2 y) = c_1 A(x) + c_2 A(y)$ for all vectors $x, y \in \mathbb{R}^k$ and all scalars $c_1, c_2 \in \mathbb{R}$,
and one often writes Ax instead of $A(x)$ due to linearity.
In this case, such a linear transformation is unique and
*it is called the (**Fréchet**) **derivative** of f at x and is denoted by $Df(x)$ or $D_x f$.*

One can define $Df(x)$ to be the linear transformation such that $f(x+h) = f(x) + Df(x)h + o(h)$,
*where $o(h)$ is a vector-valued function such that $\lim_{h \to 0, h \in \mathbb{R}^k} \frac{o(h)}{\|h\|} = 0$, called in **Landau notation**.*
Once when we fix bases of $\mathbb{R}^k$ and $\mathbb{R}^m$,
the linear transformation $Df(x)$ can be represented by a unique $m \times k$ matrix
and the vector $h \in \mathbb{R}^k$ is regarded as an $k \times 1$ matrix;
we denote this matrix by $Df(x)$ without ambiguity
*and call it the **Jacobian matrix** of f at x (depending on the given bases).*
If $m = k$,
*the function $J_f : S \to \mathbb{R}$ given by $x \mapsto det(Df(x))$ is called the **Jacobian** of f,*
which dose not depend on the basis used to construct the Jacobian matrix. .

Proposition 26.2.
Let us fix bases of $\mathbb{R}^k$ and $\mathbb{R}^m$, and
let $S \subset \mathbb{R}^k$ be open, $f : S \to \mathbb{R}^m$ with $f = (f_1, \cdots, f_m)^T$ written as a column in $\mathbb{R}^m$, and $a \in S$.
Then f is differentiable at $a \iff$ each of its components $f_1, \cdots, f_m$ is differentiable at a.
In this case, $Df(a) = \begin{bmatrix} \triangledown f_1(a) \\ \vdots \\ \triangledown f_m(a) \end{bmatrix} = \begin{bmatrix} \partial_1 f_1(a) & \cdots & \partial_k f_1(a) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(a) & \cdots & \partial_k f_m(a) \end{bmatrix}$.

Theorem 26.3 (chain rule for vector-valued functions)**.**
Let $g : \mathbb{R}^k \to \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}^k$ and let $f : \mathbb{R}^m \to \mathbb{R}^n$ be differentiable at $g(a)$.
Then the composition $f \circ g : \mathbb{R}^k \to \mathbb{R}^n$ is differentiable at a and $D(f \circ g)(a) = Df(g(a))Dg(a)$.

Definition 26.4. (*directional derivative of a vector-valued function*)
Let $f : \mathbb{R}^k \to \mathbb{R}^m$ be a function and $a \in \mathbb{R}^k$ be a point.
*Let u be a **unit vector** in $\mathbb{R}^k$, that is, $\|u\| = 1$.*
*The **directional derivative, or Gâteaux derivative, of** f **at** a, **in the direction of** u is defined to be*

$$\partial_u f(a) = \frac{d}{dt} f(a + tu)\Big|_{t=0} = \lim_{t \to 0} \frac{f(a+tu) - f(a)}{t} \text{ provided the limit exists (as a vector in } \mathbb{R}^m).$$

Theorem 26.5.
Differentiability of implies the existence of all directional derivatives and $Df(a)u = \partial_u f(a)$,
where $f : \mathbb{R}^k \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^k$ and $u \in \mathbb{R}^k$ is a unit vector written as a column vector.

Remark 26.6.
Let $L(\mathbb{R}^k, \mathbb{R}^m)$ be the set of all linear transformation from the vector space $\mathbb{R}^k$ to the vector space $\mathbb{R}^m$.
*For $A \in L(\mathbb{R}^k, \mathbb{R}^m)$, define $\|A\| = \sup\{\frac{\|Ax\|}{\|x\|} : x \in \mathbb{R}^k, x \neq 0\}$, called the **operator norm** or **induced norm** on $L(\mathbb{R}^k, \mathbb{R}^m)$*
and for $A, B \in L(\mathbb{R}^k, \mathbb{R}^m)$, define $d(A, B) = \|A - B\|$.
Then $L(\mathbb{R}^k, \mathbb{R}^m)$ with d is a metric space.
If S is a metric space, a_{ij} is a real-valued continuous function on S for all $1 \leq i \leq m$ and $1 \leq j \leq k$, and
for each $z \in S$, $A(z) \in L(\mathbb{R}^k, \mathbb{R}^m)$ has a matrix representation with entries $a_{ij}(z)$ while the bases of $\mathbb{R}^k$ and $\mathbb{R}^m$ are fixed,
then the function $z \mapsto A(z)$ is a continuous function from S to $L(\mathbb{R}^k, \mathbb{R}^m)$.

Theorem 26.7. (C^1 *functions*)
Let $S \subset \mathbb{R}^k$ be open and $f : S \to \mathbb{R}^m$ with $f = (f_1, \cdots, f_m)$.
Then the partial derivatives $\partial_j f_i$ all exist and are continuous on S
*$\iff f$ is **continuously differentiable** on S,*
that is, f is differentiable on S and $Df : S \to L(\mathbb{R}^k, \mathbb{R}^m)$ is continuous on S
(in this case we say that f is C^1 on S or $f \in C^1(S)$). (advanced exercise)