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

Definition 21.1. (C^r *and* C^∞ *functions*)
Let $U \subseteq \mathbb{R}^k$ *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 $S \subseteq \mathbb{R}^k$ *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 $S \subseteq \mathbb{R}^k$ *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 $S \subseteq \mathbb{R}^2$ *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 $\alpha \in \mathbb{N}_0^k \subseteq \mathbb{R}^k$ *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 $S \subseteq \mathbb{R}^k$ *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 $I \subseteq \mathbb{R}$ is an open interval and $m \geq 0$ is an integer.
Then for any $x \in I$,

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

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 $I \subseteq \mathbb{R}$ is an open interval and $m \geq 1$ is an integer.
Then for any $x \in I$,

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

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 $I \subseteq \mathbb{R}$ is an open interval and $m \geq 0$ 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 is C^∞ on I, that is, 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 we call the series $\sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n$ the **Taylor series** for f about c at $x \in I$;*
which, in general, needs NOT converge to the function $f(x)$, refer to Example 22.6 and Corollary 69.2.
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* $\mathbb{R}$.

Definition 22.5.
Let $f : I \to \mathbb{R}$ be a C^∞ function, where $I \subseteq \mathbb{R}$ is an open interval.
*We say that f is **analytic** on I*
if for any $c \in I$, there exists a $\delta > 0$ such that $(c-\delta, c+\delta) \subseteq I$ and for any $x \in (c-\delta, c+\delta)$, one has $f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n$.

Example 22.6.
Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \begin{cases} e^{-1/x}, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$
Then f is C^∞ on $\mathbb{R}$, analytic on $\mathbb{R}\backslash\{0\}$, and is not analytic at the origin.

23 Taylor's theorem in several variables

Theorem 23.1 (Taylor's theorem in several variables).
Let $S \subseteq \mathbb{R}^k$ 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 $S \subseteq \mathbb{R}^k$ 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 $\lim\limits_{x \to c} \dfrac{E(x-c)}{\|x-c\|^m} = 0$.
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)

Recall 24.2. (*the critical point theorem in several variables*)
Let $S \subseteq \mathbb{R}^k$ 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., $\triangledown f(c) = 0$.*

Theorem 24.3 (the second derivative test in several variables).
Let $S \subseteq \mathbb{R}^k$ be open, $f : S \to \mathbb{R}$ be C^2 at $c \in S$, and c is a critical point of f.
*Let $H_f(c) = \begin{bmatrix} \partial_{11}^2 f(c) & \cdots & \partial_{1k}^2 f(c) \\ \vdots & \ddots & \vdots \\ \partial_{k1}^2 f(c) & \cdots & \partial_{kk}^2 f(c) \end{bmatrix}$, 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 $S \subseteq \mathbb{R}^2$ 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, $H_f(c) = \begin{bmatrix} \partial_{11}^2 f(c) & \partial_{12}^2 f(c) \\ \partial_{21}^2 f(c) & \partial_{22}^2 f(c) \end{bmatrix}$.*

1. *If $det(H_f(c)) > 0$ and $\partial_{11}^2(c) > 0$, then f has a local minimum at c.*
2. *If $det(H_f(c)) > 0$ and $\partial_{11}^2(c) < 0$, 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 $S \subseteq \mathbb{R}^k$ 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 $S \subseteq \mathbb{R}^k$ 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 $g = (g_1, \ldots, g_m)$.
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 $\triangledown g_1(a), \ldots, \triangledown g_m(a)$ are linearly independent ($\triangledown g_1(a) \neq 0$ for $m = 1$),
then there exists m real numbers $\lambda_1, \ldots, \lambda_m$ such that

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

*Here the numbers $\lambda_1, \ldots, \lambda_m$ 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 $S \subseteq \mathbb{R}^k$ 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 $\lim_{h \to 0, h \in \mathbb{R}^k} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0$*;*
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 Df_x.*

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 $\mathbb{R}^m$ and write vectors in column form,
the linear transformation $Df(x)$ can be represented by a unique $m \times k$ matrix, denoted by $Df(x)$ without ambiguity,
the value $Df(x)h$ of the linear transformation becomes the matrix $Df(x)$ acting on the column vector h,
moreover, the column vector $h \in \mathbb{R}^k$ is regarded as an $k \times 1$ matrix,
and the action $Df(x)h$ can also be treated as a multiplication of matrices;
*we call $Df(x)$ the **Jacobian matrix** of f at x (depending on the given bases).*
If $m = k$, then the Jacobian matrix $Df(x)$ is a $k \times k$ matrix,
*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 bases used to construct the Jacobian matrix.

Proposition 26.2.
Let us fix bases of $\mathbb{R}^k$ and $\mathbb{R}^m$, and
let $S \subseteq \mathbb{R}^k$ 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, $Df(a) = \begin{bmatrix} \triangledown f_1(a) \\ \vdots \\ \triangledown f_m(a) \end{bmatrix} = \begin{bmatrix} \partial_1 f_1(a) & \cdots & \partial_k f_1(a) \\ \vdots & \ddots & \vdots \\ \partial_1 f_m(a) & \cdots & \partial_k f_m(a) \end{bmatrix}$.
If $m = 1$, then $Df(a) = \triangledown f(a)$ and $Df(a)h^T = \triangledown f(a) \cdot h$, while $\triangledown f$ and h were written in row form before.
in some textbook, vectors and gradients are written in column form then $Df(a)^T = \triangledown f(a)$ and $Df(a)h = \triangledown f(a) \cdot h$.

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 (**two-sided**) **directional derivative, or Gâteaux derivative, of** f **at** a, **in the direction of** u is defined to be*

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

If $k = 1$, then the one-sided directional derivative with fit well with one-side derivative.

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 $S \subseteq \mathbb{R}^k$ 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)

27 The mean value theorem and the implicit function theorem

Theorem 27.1 (the mean value theorem for vector-valued functions).
Let $S \subseteq \mathbb{R}^k$ be open and let $f : S \to \mathbb{R}^m$ a function.

1. *If $a, b \in S$ such that $L := \{a + t(b-a) : 0 \le t \le 1\} \subseteq S$,
and if f is continuous on L and is differentiable on $L \backslash \{a, b\}$,
then for any vector $u \in \mathbb{R}^m$ written as a column vector,
there exists $c \in L$ such that $u \cdot (f(b) - f(a)) = u \cdot (Df(c)(b-a))$, where $b - a$ is written as a column vector in $\mathbb{R}^k$.
In particular, if $m = 1$, then there exists $c \in L$ such that $f(b) - f(a) = \triangledown f(c) \cdot (b-a)^T$,
where $\triangledown f(c) := (\partial_1 f(c), \cdots, \partial_k f(c))$ and $(b-a)^T$ is rewritten as a row vector, identical to Theorem 19.1.*

2. *If S is convex, f is differentiable on S,
and there exists $M \in [0, \infty)$ such that $\|Df(x)\| \le M$ for all $x \in S$,
where $\|Df(x)\| := \max\{\|Df(x)y\| : y \in \mathbb{R}^k, \|y\| = 1\} = \sup\{\frac{\|Df(x)y\|}{\|y\|} : y \in \mathbb{R}^k, y \ne 0\}$, called the **norm** of $Df(x)$,
then $\|f(b) - f(a)\| \le M\|b - a\|$ for all $a, b \in S$.
(Note that the usual mean value theorem does not hold, e.g., the function $t \mapsto (\cos(t), \sin(t))$ on $[0, 2\pi]$.)*

Theorem 27.2 (the implicit function theorem in one variable).
*Let $f : \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}$ be C^1 on some neighborhood of $(a, b) \in \mathbb{R}^k \times \mathbb{R}$
and write $(x, y) \mapsto f(x, y)$.
If $f(a, b) = 0$ and $\partial_{k+1} f(a, b) \ne 0$,
then there exist $r_0 > 0$ and $r_1 > 0$ such that
for any $x \in B_{r_0}(a)$ there exists a unique $y \in B_{r_1}(b)$ such that $f(x, y) = 0$,
where $B_{r_0}(a) = \{x \in \mathbb{R}^k : \|x - a\| < r_0\}$ and $B_{r_1}(b) = \{y \in \mathbb{R} : |y - b| < r_1\}$);
denote this y by $g(x)$, then $g : B_{r_0}(a) \to B_{r_1}(b)$ is a C^1 function on $B_{r_0}(a)$ with $g(a) = b$
and for all $x \in B_{r_0}(a)$, $f(x, g(x)) = 0$ and $\partial_i g(x) = -\dfrac{\partial_i f(x, g(x))}{\partial_{k+1} f(x, g(x))}$ for all $1 \le i \le k$.*

Corollary 27.3.
*Let $f : \mathbb{R}^k \to \mathbb{R}$ be C^1 on $\mathbb{R}^k$ and let $\gamma = \{x \in \mathbb{R}^k : f(x) = 0\}$.
If $a \in \gamma$ with $\triangledown f(a) \ne 0$,
then there exists a neighborhood U of a in $\mathbb{R}^k$ such that $\gamma \cap U$ is the graph of a C^1 function.*

Theorem 27.4 (the implicit function theorem in several variables).
*Let $f : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m$ be C^1 on some neighborhood of $(a, b) \in \mathbb{R}^k \times \mathbb{R}^m$
and write $(x, y) \mapsto f(x, y)$ and $f = (f_1, \cdots, f_m)$.
If $f(a, b) = 0$ and $det(A) \ne 0$, where $A = \begin{bmatrix} \partial_{k+1} f_1 & \cdots & \partial_{k+m} f_1 \\ \vdots & \ddots & \vdots \\ \partial_{k+1} f_m & \cdots & \partial_{k+m} f_m \end{bmatrix} (a, b)$,
then there exist $r_0 > 0$ and $r_1 > 0$ such that
for any $x \in B_{r_0}(a)$ there exists a unique $y \in B_{r_1}(b)$ such that $f(x, y) = 0$,
where $B_{r_0}(a) = \{x \in \mathbb{R}^k : \|x - a\| < r_0\}$ and $B_{r_1}(b) = \{y \in \mathbb{R}^m : \|y - b\| < r_1\}$;
denote this y by $g(x)$, then $g : B_{r_0}(a) \to B_{r_1}(b)$ is a C^1 function on $B_{r_0}(a)$ with $g(a) = b$
and for all $x \in B_{r_0}(a)$, $f(x, g(x)) = 0$ and $\partial_i g(x)$ for $1 \le i \le k$ can be computed
by differentiating $f(x, g(x)) = 0$ with respect to x_i
and solving the resulting linear system of equations for $\partial_i g_1, \cdots, \partial_i g_m$, where $g = (g_1, \cdots, g_m)$ as follows:*

$$\begin{bmatrix} \partial_1 f_1 & \cdots & \partial_k f_1 \\ \vdots & \ddots & \vdots \\ \partial_1 f_m & \cdots & \partial_k f_m \end{bmatrix} + \begin{bmatrix} \partial_{k+1} f_1 & \cdots & \partial_{k+m} f_1 \\ \vdots & \ddots & \vdots \\ \partial_{k+1} f_m & \cdots & \partial_{k+m} f_m \end{bmatrix} \begin{bmatrix} \partial_1 g_1 & \cdots & \partial_k g_1 \\ \vdots & \ddots & \vdots \\ \partial_1 g_m & \cdots & \partial_k g_m \end{bmatrix} = 0.$$

28 Curves in the plane and surfaces in the space

Definition 28.1 (C^1 curve and surface)**.**

1. *A set $\gamma \subseteq \mathbb{R}^2$ is called a C^1* ***curve*** *if every $a \in \gamma$ has a neighborhood U such that $\gamma \cap U$ is the graph of a C^1 function g (either $y = g(x)$ or $x = g(y)$).*

2. *A set $\gamma \subseteq \mathbb{R}^3$ is called a C^1* ***surface*** *if every $a \in \gamma$ has a neighborhood U such that $\gamma \cap U$ is the graph of a C^1 function g (either $z = g(x,y)$, $y = g(x,z)$, or $x = g(y,z)$).*

Theorem 28.2 (sufficient conditions for C^1 curves in $\mathbb{R}^2$)**.**

1. *(a non-parametric form)*
   Let $S \subseteq \mathbb{R}^2$ be open, $f : S \to \mathbb{R}$ be C^1 on S, and $\gamma = \{(x,y) \in S : f(x,y) = 0\}$.
   If $a \in \gamma$ and $\nabla f(a) \neq 0$,
   then there exists a neighborhood U of a in $\mathbb{R}^2$
   such that $\gamma \cap U$ is the graph of a C^1 function g (either $y = g(x)$ or $x = g(y)$).

2. *(a parametric form)*
   Let $f : (a,b) \to \mathbb{R}^2$ be C^1 on (a,b).
   If $t_0 \in (a,b)$ and $Df(t_0) = f'(t_0) \neq 0$,
   then there exists an open interval I of t_0 in (a,b)
   such that the set $\{f(t) : t \in I\}$ is the graph of a C^1 function g
   (either $y = g(x)$ or $x = g(y)$).

Theorem 28.3 (sufficient conditions for C^1 surface in $\mathbb{R}^3$)**.**

1. *(a non-parametric form)*
   Let $S \subseteq \mathbb{R}^3$ be open, $f : S \to \mathbb{R}$ be C^1 on S
   which is denoted by $(x,y,z) \mapsto f(x,y,z)$, and let $\gamma = \{(x,y,z) \in S : f(x,y,z) = 0\}$.
   If $a \in \gamma$ and $\nabla f(a) \neq 0$,
   then there exists a neighborhood U of a in $\mathbb{R}^3$
   such that $\gamma \cap U$ is the graph of a C^1 function g
   (either $z = g(x,y)$, $y = g(x,z)$, or $x = g(y,z)$).

2. *(a parametric form)*
   Let $S \subseteq \mathbb{R}^2$ be open and $f : S \to \mathbb{R}^3$ be C^1 on S, denoted by $(s,t) \mapsto f(s,t)$.
   If $(s_0,t_0) \in S$ and the vectors $\frac{\partial f}{\partial s}(s_0,t_0)$ and $\frac{\partial f}{\partial t}(s_0,t_0)$ are linearly independent
   (equivalently, if $(s_0,t_0) \in S$ and the cross product $\frac{\partial f}{\partial s}(s_0,t_0) \times \frac{\partial f}{\partial t}(s_0,t_0) \neq 0$),
   then there exists a neighborhood U of (s_0,t_0) in S
   such that the set $\{f(s,t) : (s,t) \in U\}$ is the graph of a C^1 function g
   (either $z = g(x,y)$, $y = g(x,z)$, or $x = g(y,z)$)..

29 Curves in $\mathbb{R}^3$ and n-dimensional manifolds in $\mathbb{R}^k$

Definition 29.1. (*curves in* $\mathbb{R}^3$)
A C^1 ***curve*** γ *in* $\mathbb{R}^3$ *is defined to be one of the following:*

1. (*the non-parametric form*) *the set* $\gamma = \{x \in U : f(x) = 0\}$,
   where $U \subseteq \mathbb{R}^3$ *be open and* $f : U \to \mathbb{R}^2$ *is a* C^1 *function with* $f = (f_1, f_2)$
   such that the vectors $\triangledown f_1(x)$ *and* $\triangledown f_2(x)$ *are* ***linearly independent*** *at each* $x \in \gamma$,
   or equivalently, the matrix $Df(x)$ *has rank* 2 *at every* $x \in \gamma$.

2. (*the parametric form*) *the set* $\gamma = \{f(t) : t \in V\}$,
   where $V \subseteq \mathbb{R}$ *is open and* $f : V \to \mathbb{R}^3$ *is a* C^1 *function*
   such that $f'(t) \neq 0$ *at each* $t \in V$.

Definition 29.2. (*manifold in* $\mathbb{R}^k$)
A C^1 n***-dimensional manifold*** γ *in* $\mathbb{R}^k$ *with* $n < k$ *is defined to be one of the following:*

1. (*the non-parametric form*) *the set* $\gamma = \{x \in U : f(x) = 0\}$,
   where $U \subseteq \mathbb{R}^k$ *be open and* $f : U \to \mathbb{R}^{k-n}$ *is a* C^1 *function with* $f = (f_1, \ldots, f_{k-n})$
   such that the vectors $\triangledown f_1(x), \ldots, \triangledown f_{k-n}(x)$ *are* ***linearly independent*** *at each* $x \in \gamma$,
   or equivalently, the matrix $Df(x)$ *has rank* $k - n$ *at every* $x \in \gamma$.

2. (*the parametric form*) *the set* $\gamma = \{f(t) : t \in V\}$,
   where $V \subseteq \mathbb{R}^n$ *is open and* $f : V \to \mathbb{R}^k$ *is a* C^1 *function*
   such that $\frac{\partial f(t)}{\partial t_i}$, $i = 1, \ldots, n$, *are* ***linearly independent*** *at each* $t \in V$,
   or equivalently, the matrix $Df(t)$ *has rank* n *at each* $t \in V$.

In fact, in the theory of Differential Geometry,
we define a C^1 n***-dimensional manifold*** γ *by using local charts so that* γ *is locally like* $\mathbb{R}^n$,
then the above becomes sufficient conditions for γ *being a* C^1 n*-dimensional manifold*
and the tangent space of γ *at each point on it is well defined.*

30 The inverse function theorem and the higher order Fréchet derivative

Theorem 30.1 (the inverse function theorem).
Let $U, V \subseteq \mathbb{R}^k$ be open and $f : U \to V$ be C^1 on U.
If $a \in U$ and the Jacobian $det(Df(a)) \neq 0$,
then there exist open neighborhoods $U_1 \subseteq U$ of a and $V_1 \subseteq V$ of $f(a)$
such that f is one-to-one from U_1 onto V_1,
the inverse function $f^{-1} : V_1 \to U_1$ is C^1,
and $Df^{-1}(f(x)) = [Df(x)]^{-1}$ for all $x \in U_1$.

Remark 30.2. (*globally one-to-one*)
Let $U, V \subseteq \mathbb{R}^k$ be open. Suppose $f : U \to V$ is C^1 and $det(Df(x)) \neq 0$ for all $x \in U$.
Is f one-to-one on U?

1. *If $k = 1$, the answer is YES.*
2. *If $k > 1$, the answer is NO.*

Problem 30.3 (Jacobian conjecture).
Let $f : \mathbb{R}^k \to \mathbb{R}^k$ be a map whose component functions are all polynomials.
Suppose $det(Df(x)) = 1$ for all $x \in \mathbb{R}^k$.
Is f one-to-one on $\mathbb{R}^k$?
(If yes, then one can prove that the inverse of f is a map defined on $\mathbb{R}^k$
whose component functions are also all polynomials.)
The conjecture that the answer is yes is called ***Jacobian conjecture****;*
this is a famous unsolved problem.

Remark 30.4. (*higher order Fréchet derivative*)
The Fréchet derivative of a function $f : U \subseteq \mathbb{R}^k \to \mathbb{R}^m$ gives a function $Df : U \to L(\mathbb{R}^k, \mathbb{R}^m)$,
that is, $Df(x) \in L(\mathbb{R}^k, \mathbb{R}^m))$ such that $\lim_{h\to 0} \frac{\|f(x+h)-f(x)-Df(x)h\|_{\mathbb{R}^m}}{\|h\|_{\mathbb{R}^k}} = 0$,
where $L(\mathbb{R}^k, \mathbb{R}^m)$ is the set of all linear transformations from $\mathbb{R}^k$ to $\mathbb{R}^m$ with the metric induced from the operator norm,
which is isomorphic to the Euclidean space $\mathbb{R}^{km}$.
Similarly, the Fréchet derivative of the function Df gives a function $D(Df) : U \to L(\mathbb{R}^k, L(\mathbb{R}^k, \mathbb{R}^m))$,
called the ***second derivation*** *of f, denoted by $x \mapsto D^2 f_x := D(Df)(x)$, that is, $D(Df(x) \in L(\mathbb{R}^k, L(\mathbb{R}^k, \mathbb{R}^m))$ such that*

$$\lim_{h\to 0} \frac{\|Df(x+h) - Df(x) - D(Df)(x)(h)\|_{L(\mathbb{R}^k,\mathbb{R}^m)}}{\|h\|_{\mathbb{R}^k}} = 0,$$

where $L(\mathbb{R}^k, L(\mathbb{R}^k, \mathbb{R}^m))$ is isomorphic to the set of maps in $L(\mathbb{R}^k \times \mathbb{R}^k, \mathbb{R}^m)$ which are linear in each factor of $\mathbb{R}^k$ separately,
called the set of all ***bilinear maps*** *from $\mathbb{R}^k$ to $\mathbb{R}^m$ and denote by $L^2(\mathbb{R}^k, \mathbb{R}^m)$, and so we write $D^2 f_x(h, v) = D^2 f_x(h)(v)$.*

Assume that f is C^2 on U. Then $\partial^2_{ij} f(x) = \partial^2_{ji} f(x)$ for all $1 \le i, j \le k,$, which is a vector in $\mathbb{R}^m$.
Let $\{e^i : 1 \le i \le k\}$ denote the standard basis of $\mathbb{R}^k$. Then for vectors $h = \sum_{i=1}^k h_i e^i$ and $v = \sum_{i=1}^k v_i e^i$, we have

$$D^2 f_x(h, v) = \sum_{1\le i,j\le k} \partial^2_{ij} f(x) h_i v_j = \sum_{1\le i,j\le k} \partial^2_{ji} f(x) v_j h_i = D^2 f_x(v, h)$$

,
which makes $D^2 f_x$ a ***symmetric bilinear*** *function from $\mathbb{R}^k \times \mathbb{R}^k$ to $\mathbb{R}^m$.*
Similarly, one can define ***the r-derivative of f at x,*** *which is a symmetric r-linear form, $D^r f_x \in L^r(\mathbb{R}^k, \mathbb{R}^m)$.*
Moreover, one has ***the chain rule***

$$D^2(g \circ f)_x(h, v) = D^2 g_{f(x)}(Df_x h, Df_x v) + Dg_{f(x)} D^2 f_x(h, v),$$

and ***the Taylor theorem*** *for C^r function $f : U \subseteq \mathbb{R}^k \to \mathbb{R}^m$,*

$$f(x) = \sum_{n=0}^{r} \frac{D^n f_c(x - c, \cdots, x - c)}{n!} + R_{f,c,r}(x - c),$$

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

Assume that $m = 1$, that is, the codomain is the base field.
Then $D^2 f_x$ is a symmetric bilinear form with the $k \times k$ matrix $\left(\partial^2_{ji} f(x)\right)_{1\le,j\le k}$, that is, the Hessian matrix of f at x,
and it gives a quadratic form $D^2 f_c(x - c, x - c) = (x - c)^T D^2 f_c(x - c) = \sum_{1\le i,j\le k} \partial^2_{ji} f(x)(x_j - c_j)(x_i - c_i)$.