Chapter 6. Higher Order Derivatives

6.1 Theorem: *(Iterated Limits) Let I and J be open intervals in $\mathbb{R}$ with $a \in I$ and $b \in J$, let $U = (I \times J) \setminus \{(a,b)\}$, and let $f : U \to \mathbb{R}$. Suppose that $\lim\limits_{y \to b} f(x,y)$ exists for every $x \in I$ and that $\lim\limits_{(x,y)\to(a,b)} f(x,y) = u \in \mathbb{R}$. Then $\lim\limits_{x \to a} \lim\limits_{t \to b} f(x,y) = u$.*

Proof: Define $g : I \to \mathbb{R}$ by $g(x) = \lim\limits_{y \to b} f(x,y)$. Let $\epsilon > 0$. Since $\lim\limits_{(x,y)\to(a,b)} f(x,y) = u$ we can choose $\delta > 0$ such that for all $(x,y) \in U$ with $0 < \big|(x,y) - (a,b)\big| \leq 2\delta$ we have $\big|f(x,y) - u\big| \leq \epsilon$. Let $x \in I$ with $0 < |x-a| \leq \delta$. For all $y \in J$ with $0 < |y-b| \leq \delta$ we have $0 < \big|(x,y) - (a,b)\big| \leq |x-a| + |y-b| \leq 2\delta$ and so $\big|f(x,y) - u\big| \leq \epsilon$ and hence

$$|g(x) - u| \leq \big|g(x) - f(x,y)\big| + \big|f(x,y) - u\big| \leq \big|g(x) - f(x,y)\big| + \epsilon.$$

Take the limit as $y \to b$ on both sides to get $|g(x) - u| \leq \epsilon$. Thus $\lim\limits_{x \to a} g(x) = u$, as required.

6.2 Theorem: *(Mixed Partials Commute) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is open in $\mathbb{R}^n$ with $a \in U$, and let $k, \ell \in \{1, \cdots, n\}$. Suppose $\frac{\partial^2 f}{\partial x_k \partial x_\ell}(x)$ exists in U and is continuous at a, $\frac{\partial f}{\partial x_k}(x)$ exists and is continuous in U, and $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(a)$ exists. Then $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(a) = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(a)$.*

Proof: When $k = \ell$ there is nothing to prove, so suppose that $k \neq \ell$. Choose $r > 0$ so that $B(a, 2r) \subseteq U$. For $|x| < r$ and $|y| < r$ note that the points a, $a + xe_k$, $a + ye_\ell$ and $a + xe_k + ye_\ell$ all lie in $B(a, 2r)$. For $|X| < r$ and $|y| < r$, define

$$g(x,y) = f(a + xe_k + ye_\ell) - f(a + xe_k) - f(a + ye_\ell) + f(a).$$

By the Mean Value Theorem, applied to the function $f(a + xe_k + ye_\ell) - f(a + ye_\ell)$ as a function of y, we can choose t between 0 and y such that

$$y\big(\tfrac{\partial f}{\partial x_\ell}(a + xe_k + te_\ell) - \tfrac{\partial f}{\partial x_\ell}(a + te_\ell)\big) = g(x,y).$$

By the Mean Value Theorem, applied to the function $\frac{\partial f}{\partial x_\ell}(a + xe_k + te_\ell)$ as a function of x, we can choose s between 0 and x such that

$$x\, \tfrac{\partial^2 f}{\partial x_k \partial x_\ell}(a + se_k + te_\ell) = \tfrac{\partial f}{\partial x_\ell}(a + xe_k + te_\ell) - \tfrac{\partial f}{\partial x_\ell}(a + te_\ell).$$

Also by the Mean Value Theorem, applied to the function $f(a + xe_k + ye_\ell) - f(a + xe_k)$ as a function of x, we can choose r between 0 and x such that

$$x\big(\tfrac{\partial f}{\partial x_k}(a + re_k + ye_\ell) - \tfrac{\partial f}{\partial x_k}(a + re_\ell)\big) = g(x,y).$$

Then for $|x| < r$ and $0 < |y| < r$ we have

$$\frac{\frac{\partial f}{\partial x_k}(a + re_k + ye_\ell) - \frac{\partial f}{\partial x_k}(a + re_k)}{y} = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(a + se_k + te_\ell).$$

Since $\frac{\partial^2 f}{\partial x_k \partial x_\ell}$ is continuous, the limit on the right as $(x,y) \to (0,0)$ is equal to $\frac{\partial^2 f}{\partial x_k \partial x_\ell}(a)$, and since $\frac{\partial f}{\partial x_k}$ is continuous, the limit as $y \to 0$ of the limit as $x \to 0$ on the left is equal to $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(a)$, so the desired result follows from the above lemma.

6.3 Corollary: *If $U \subseteq \mathbb{R}^n$ is open and $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$ in U then we have $\frac{\partial^2 f}{\partial x_\ell \partial x_k}(x) = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(x)$ for all $x \in U$ and for all k, ℓ.*

6.4 Exercise: Verify that for $f(x,y) = \frac{x^2}{x^2+y^2}$ we have $\lim\limits_{x \to 0} \lim\limits_{y \to 0} f(x,y) \neq \lim\limits_{y \to 0} \lim\limits_{x \to 0} f(x,y)$.

6.5 Exercise: Let $f(x,y) = \left\{ \begin{array}{cl} \dfrac{xy(x^2-y^2)}{x^2+y^2} & \text{, if } (x,y) \neq (0,0) \\ 0 & \text{, if } (x,y) = (0,0) \end{array} \right\}$. Verify that the mixed partial derivatives $\frac{\partial^2 f}{\partial x \partial y}(0,0)$ and $\frac{\partial^2 f}{\partial y \partial x}(0,0)$ both exist, but they are not equal.

6.6 Definition: for $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, where U is open in $\mathbb{R}^n$ with $a \in U$, we define $D^0 f(a) = f(a)$ and for $\ell \in \mathbb{Z}^+$ we define the ℓ^{th} **total differential** of f at a to be the map $D^\ell f(a) : \mathbb{R}^n \to \mathbb{R}$ given by

$$D^\ell f(a)(u) = \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \cdots \sum_{k_\ell=1}^{n} \frac{\partial^\ell f}{\partial x_{k_1} \partial x_{k_2} \cdots \partial x_{k_\ell}}(a)\, u_{k_1} u_{k_2} \cdots u_{k_\ell}$$

provided that all of the ℓ^{th} order partial derivatives exist at a.

6.7 Example: When $f : U \subseteq \mathbb{R}^2 \to \mathbb{R}$ is $\mathcal{C}^2$ (so the mixed partial derivatives commute) we have

$$\begin{aligned} D^0 f(u,v) &= f(a,b) \\ D^1 f(a,b)(u,v) &= \tfrac{\partial f}{\partial x}(a,b)\, u + \tfrac{\partial f}{\partial y}(a,b)\, v \\ D^2 f(a,b)(u,v) &= \tfrac{\partial f}{\partial x^2}(a,b)\, u^2 + 2 \tfrac{\partial f}{\partial x \partial y}(a,b)\, uv + \tfrac{\partial f}{\partial y^2}(a,b)\, v^2. \end{aligned}$$

6.8 Theorem: *(Taylor's Theorem) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is open in $\mathbb{R}^n$. Suppose that the m^{th} oder partial derivatives of f all exist in U. Then for all $a, x \in U$ such that $[a,x] \subseteq U$ there exists $c \in [a,x]$ such that*

$$f(x) = \sum_{\ell=0}^{m-1} \tfrac{1}{\ell!}\, D^\ell f(a)(x-a) \;+\; \tfrac{1}{m!}\, D^m f(c)(x-a).$$

Proof: Let $a, x \in U$ with $[a,x] \subseteq U$. Let $\alpha(t) = a + t(x-a)$ for all $t \in \mathbb{R}$ and note that $\alpha(t) \in U$ for $0 \leq t \leq 1$. Since U is open and α is continuous, we can choose $\delta > 0$ so that $\alpha(t) \in U$ for all $t \in I = (-\delta, 1+\delta)$. Define $g : I \to \mathbb{R}$ by $g(t) = f(\alpha(t))$. By the Chain Rule, we have

$$g'(t) = Df\big(\alpha(t)\big)\alpha'(t) = Df\big(\alpha(t)\big)(x-a) = \sum_{i=1}^{n} \tfrac{\partial f}{\partial x_i}\big(\alpha(t)\big)(x_i - a_i) = D^1 f\big(\alpha(t)\big)(x-a).$$

By the Chain Rule again, we have

$$g''(t) = \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \tfrac{\partial^2 f}{\partial x_j \partial x_i}\big(\alpha(t)\big)(x_j - a_j) \Big)(x_i - a_i) = D^2 f\big(\alpha(t)\big)(x-a).$$

An induction argument shows that

$$g^{(\ell)}(t) = D^\ell f\big(\alpha(t)\big)(x-a).$$

By Taylor's Theorem, applied to the function $g(t)$ on the interval $[0,1]$, we can choose $s \in [0,1]$ such that $g(1) = \sum_{\ell=0}^{m-1} \frac{1}{\ell!} g^{(\ell)}(0) \;+\; \frac{1}{m!} g^{(m)}(s)$, that is

$$f(x) = \sum_{\ell=0}^{m-1} \tfrac{1}{\ell!} D^\ell f(a)(x-a) \;+\; \tfrac{1}{m!} D^m f\big(\alpha(s)\big)(x-a).$$

Thus we can choose $c = \alpha(s) \in [a,x]$.

6.9 Definition: For $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, where U is open in $\mathbb{R}^n$ with $a \in U$, we define the m^{th} **Taylor polynomial** of f at a to be the polynomial

$$T^m f(a)(x) = \sum_{\ell=0}^{m} \tfrac{1}{\ell!} D^\ell f(a)(x-a)$$

provided that all the m^{th} order partial derivatives exist at a. When f is $\mathcal{C}^2$ in U (so that the mixed partial derivatives commute) we have

$$T^2 f(a)(x) = f(a) + Df(a)(x-a) + \tfrac{1}{2}\,(x-a)^T Hf(a)\,(x-a)$$

where $Hf(a) \in M_{n\times n}(\mathbb{R})$ is the symmetric matrix with entries $Hf(a)_{k,\ell} = \frac{\partial^2 f}{\partial x_k \partial x_\ell}(a)$. The matrix $Hf(a)$ is called the **Hessian matrix** of f at a.

6.10 Definition: Let $A \in M_n(\mathbb{R})$ be a symmetric matrix. We say that

(1) A is **positive-definite** when $u^T A u > 0$ for all $0 \neq u \in \mathbb{R}^n$,
(2) A is **negative-definite** when $u^T A u < 0$ for all $0 \neq u \in \mathbb{R}^n$, and
(3) A is **indefinite** when there exist $0 \neq u, v \in \mathbb{R}^n$ with $u^T A u > 0$ and $v^T A v < 0$.

6.11 Theorem: *(Characterization of Positive-Definiteness by Eigenvalues) Let $A \in M_n(\mathbb{R})$ be symmetric. Then*

(1) A is positive-definite if and only if all of the eigenvalues of A are positive,
(2) A is negative-definite if and only if all of the eigenvalues of A are negative, and
(3) A is indefinite if and only if A has a positive eigenvalue and a negative eigenvalue.

Proof: Suppose that A is positive definite. Let λ be an eigenvalue of A and let u be a unit eigenvector for λ. Then $\lambda = \lambda|u|^2 = \lambda(u \cdot u) = \lambda u \cdot u = Au \cdot u = u^T A u > 0$. Conversely, suppose that all of the eigenvalues of A are positive. Since A is symmetric, we can orthogonally diagonalize A. Choose a matrix $P \in M_n(\mathbb{R})$ with $P^T = P$ so that $P^T A P = D = \text{diag}(\lambda_1, \cdots, \lambda_n)$. Given $0 \neq u \in \mathbb{R}^n$, let $v = P^T u$. Note that $v \neq 0$ since P^T is invertible. Thus $u^T A u = u^T P D P^T u = v^T D v = \sum_{i=1}^{n} \lambda_i {v_i}^2 > 0$ since every $\lambda_i > 0$ and some $v_i \neq 0$. This proves Part (1). The proofs of Parts (2) and (3) are fairly similar.

6.12 Theorem: *(Characterization of Positive-Definiteness by Determinant) Let $A \in M_n(\mathbb{R})$ be symmetric. For each k with $1 \leq k \leq n$, let $A^{(k)}$ denote the upper-left $k \times k$ sub matrix of A. Then*

(1) A is positive-definite if and only if $\det(A^{(k)}) > 0$ for all k with $1 \leq k \leq n$, and
(2) A is negative-definite if and only if $(-1)^k \det(A^{(k)}) > 0$ for all k with $1 \leq k \leq n$.

Proof: Part (2) follows easily from Part (1) by noting that A is negative-definite if and only if $-A$ is positive-definite. We shall prove one direction of Part (1). Suppose that A is positive-definite. Let $1 \leq k \leq n$. Since $u^T A u > 0$ for all $0 \neq u \in \mathbb{R}^n$, we have $\begin{pmatrix} u^T & 0 \end{pmatrix} A \begin{pmatrix} u \\ 0 \end{pmatrix} = 0$, or equivalently $u^T A^{(k)} u > 0$, for all $0 \neq u \in \mathbb{R}^k$. This shows that $A^{(k)}$ is positive definite. By the previous theorem, all of the eigenvalues of $A^{(k)}$ are positive. Since $\det(A^{(k)})$ is equal to the product of its eigenvalues, we see that $\det(A^{(k)}) > 0$.

The proof of the other direction of Part (1) is more difficult. We shall omit the proof. It is often proven in a linear algebra course.

6.13 Exercise: Let $A = \begin{pmatrix} 3 & -1 & 2 \\ -1 & 2 & 1 \\ 2 & 1 & 5 \end{pmatrix}$. Determine whether A is positive-definite.

6.14 Definition: Let $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ and let $a \in A$. We say that f has a **local maximum value** at a when there exists $r > 0$ such that $f(a) \geq f(x)$ for all $x \in B_A(a,r)$. We say that f has a **local minimum value** at a when there exists $r > 0$ such that $f(a) \leq x$ for all $x \in B_A(a,r)$.

6.15 Exercise: Show that when $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is open in $\mathbb{R}^n$ with $a \in U$, if f has a local maximum or minimum value at a then either $Df(a) = 0$ or $Df(a)$ does not exist (that is one of the partial derivatives $\frac{\partial f}{\partial x_k}(a)$ does not exist).

6.16 Definition: Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ where U is open in $\mathbb{R}^n$. For $a \in U$, we say that a is a **critical point** of f when either $Df(a) = 0$ or $Df(a)$ does not exist. When $a \in U$ is a critical point of f but f does not have a local maximum or minimum value at a, we say that a is a **saddle point** of f.

6.17 Theorem: *(The Second Derivative Test) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ with U open in $\mathbb{R}^n$ and let $a \in U$. Suppose that f is $\mathcal{C}^2$ in U with $Df(a) = 0$. Then*

(1) if $Hf(a)$ is positive definite then f has a local minimum value at a,
(2) if $Hf(a)$ is negative definite then f has a local maximum value at a, and
(3) if $Hf(a)$ is indefinite then f has a saddle point at a.

Proof: Suppose that $Hf(a)$ is positive-definite. Then $\det\big(Hf(a)^{(k)}\big) > 0$ for $1 \leq k \leq n$. Since each determinant function $\det(A^{(k)})$ is continuous as a function in the entries of the matrix A, the set $V = \big\{x \in U \,\big|\, Hf(x)^{(k)} > 0 \text{ for } k = 1, 2, \cdots, n\big\}$ is open. Choose $r > 0$ so that $B(a,r) \subseteq V$. Then we have $u^T Hf(c)\, u > 0$ for all $0 \neq u \in \mathbb{R}^n$ and all $c \in B(a,r)$. Let $x \in B(a,r)$ with $x \neq a$. By Taylor's Theorem, we have

$$f(x) - f(a) - Df(a)(x-a) = (x-a)^T Hf(c)\,(x-a)$$

for some $c \in [a,x]$. Since $Df(a) = 0$ and $Hf(c)$ is positive-definite, we have $f(x) - f(a) > 0$. Thus f has a local minimum value at a. This proves Part (1) and Part (2) is similar.

Let us prove Part (3). Suppose there exists $0 \neq u \in \mathbb{R}^n$ such that $u^T Hf(a)\, u > 0$. Let $r > 0$ with $B(a,r) \subseteq U$ and scale the vector u if necessary so that $[a,u] \subseteq B(a,r)$. Let $\alpha(t) = a + tu$ and let $g(t) = f\big(\alpha(t)\big)$ for $0 \leq t \leq 1$. As in the proof of Taylor's Theorem, we have

$$g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\big(\alpha(t)\big)\, u_i = Df\big(\alpha(t)\big)\, u \text{ , and}$$
$$g''(t) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\big(\alpha(t)\big)\, u_i u_j = u^T Hf\big(\alpha(t)\big)\, u\,.$$

Since $g(0) = f(a)$, $g'(0) = Df(a)\, u = 0$ and $g''(0) = u^T Hf(a)\, u > 0$, it follows from single-variable calculus that we can choose t_0 with $0 < t_0 < 1$ so that $g(t_0) > g(0)$. When $x = \alpha(t_0)$ we have $x \in B(a,r)$ and $f(x) = f\big(\alpha(t_0)\big) = g(t_0) > g(0) = f(a)$, and so f does not have a local maximum value at a. Similarly, if there exists $0 \neq v \in \mathbb{R}^n$ such that $v^T Hf(a)\, v < 0$ then f does not have a local minimum value at a. Thus when $Hf(a)$ is indefinite, f has a saddle point at a.

6.18 Exercise: Find and classify the critical points of the following functions $f : \mathbb{R}^2 \to \mathbb{R}$.

(a) $f(x,y) = x^3 + 2xy + y^2$ (b) $f(x,y) = x^3 + 3x^2y - 6y^2$ (c) $f(x,y) = x^2 y\, e^{-x^2 - 2y^2}$