

# MAT237 Final Review Notes

Tingfeng Xia

April 13, 2019

## Contents

<b>1</b>	<b>Set Theory, Limit and Continuity</b>	<b>2</b>
1.1	Limits and Continuity . . . . .	2
1.2	Sequences and Completeness . . . . .	3
1.3	Compactness and Applications . . . . .	4
1.4	The Intermediate Value Theorem . . . . .	6
<b>2</b>	<b>Differential Calculus</b>	<b>6</b>
2.1	Differentiation of Real-Fcns . . . . .	6
2.2	Differentiation of Vec-Fcns . . . . .	8
2.3	The Chain Rule . . . . .	9
2.4	The Mean Value Theorem . . . . .	10
2.5	Higher Order Derivatives . . . . .	11
2.6	Taylor's Theorem . . . . .	11
2.7	Critical Points . . . . .	12

# 1 Set Theory, Limit and Continuity

## 1.1 Limits and Continuity

Notice that major portion of this part of note is available in a separate file<sup>1</sup>, typeset-ted using one-Note prior to the existence of these notes. There is one final (important) result in the limits part to be stated in this section.

**Theorem: (Limit Existence, HW1.1).** Consider the limit of the following flavour

$$\lim_{(x,y) \rightarrow (0,0)} \frac{|x|^{r_1} |y|^{r_2}}{|x|^{s_1} + |y|^{s_2}}$$

where  $r_1, r_2, s_1, s_2 \in \mathbb{R}^{>0}$ . Then the limit exists, and is equal to zero if  $\frac{r_1}{s_1} + \frac{r_2}{s_2} > 1$ .

**Definition: Continuity** If  $S \subseteq \mathbb{R}^n$ , then a function  $\mathbf{f} : S \rightarrow \mathbb{R}^k$  is continuous at  $\mathbf{a} \in S$  if

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \mathbf{x} \in S \wedge |\mathbf{x} - \mathbf{a}| < \delta \implies |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})| < \varepsilon$$

also, we say that  $\mathbf{f}$  is continuous if  $\mathbf{f}$  is continuous at every point in  $S$ . There is an alternative way of defining the continuity, which is as follows: if  $\mathbf{a}$  is a point where it makes sense to talk about  $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x})$ , then

$$\mathbf{f} : S \rightarrow \mathbb{R}^k \text{ is CTS at } \mathbf{a} \in S \iff \lim_{\mathbf{x} \rightarrow \mathbf{a}, \mathbf{x} \in S} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$$

**Theorem: (Properties of Continuity)** Consider  $S \subseteq \mathbb{R}^n, \mathbf{a} \in S$

1. A vector valued function  $\mathbf{f} : S \rightarrow \mathbb{R}^k$  is CTS at  $\mathbf{a}$  if and only if all of its component real valued functions are CTS at  $\mathbf{a}$ .
2. If two vector valued functions each are CTS at  $\mathbf{a}$ , then so is their sum.
3. If two real valued functions  $f, g$  each are CTS at  $\mathbf{a}$ , then so is their product. In addition if  $g(\mathbf{a}) \neq 0$ , then  $f/g$  is also CTS at  $\mathbf{a}$

---

<sup>1</sup>see first\_two.pdf available in the same folder

4. If the vectors valued functions  $\mathbf{f}$  is CTS at  $\mathbf{a}$ , and  $\mathbf{g}$  is CTS at  $\mathbf{f}(\mathbf{a})$ , then so is their composition  $\mathbf{g} \circ \mathbf{f}$ . Notice that here we are implicitly assuming the two functions are compatible to each other in composition.
5. Elementary functions, such as trigonometric functions, trigonometric inverses, polynomials, exponentials and logarithm are all continuous on their domains.

**Theorem: (Sets w/ Continuity)** Assume that  $\mathbf{f}$  is a function  $\mathbb{R}^n \rightarrow \mathbb{R}^k$ . Then, the following are equivalent

- $\mathbf{f}$  is continuous
- For every open set  $U \subseteq \mathbb{R}^k$ , the set  $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{f}(\mathbf{x}) \in U\}$  is open
- For every open set  $K \subseteq \mathbb{R}^k$ , the set  $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{f}(\mathbf{x}) \in K\}$  is closed

There is a big family of question that is based on this equivalence relationship. Consider  $S = \{\mathbf{x} \in \mathbb{R}^3 : f(\mathbf{x}) \in T\}, T \subseteq \text{domain of } f$ . Then, using the fact that  $f$  is CTS function, we can know that the set  $S$  is closed/open iff the pre-image  $T$  is closed/open.

**Recognizing Sets** The general principle is: if any set is defined using continuous functions (possibly more than one of them) and strict inequalities (i.e.,  $<$  or  $>$ ) then we instantly recognize it as open. And if any set is defined using continuous functions and non-strict inequalities (i.e.,  $\leq$  or  $\geq$ ) or equality, then we instantly recognize it as closed.

## 1.2 Sequences and Completeness

**Theorem: (Bounded Sequence Theorem)** Every bounded sequence in  $\mathbb{R}^n$  has a subsequence that converges to a limit.

**Theorem.** Let  $\{\mathbf{a}_j\}_j \subseteq \mathbb{R}^n$  be sequence, let  $\mathbf{L} = (L_1, \dots, L_n) \in \mathbb{R}^n$ , then

$$\lim_{j \rightarrow \infty} \mathbf{a}_j = \mathbf{L} \iff \lim_{j \rightarrow \infty} a_{jk} = L_k \text{ for all } k = 1, \dots, n$$

**Theorem: (Completeness Axiom)** Every bounded non-empty set of real number has a least upper bound

### Definitions: Upper Bound, Supremum

1. A number  $a$  is an upper bound for a set  $S \subseteq \mathbb{R}$  if  $a \geq x, \forall x \in S$
2. A number  $a$  is the least upper bound (supremum) for  $S$  if  $a$  is an upper bound for  $S$ , and for every upper bound  $b$  for  $S$ ,  $a \leq b$ . We usually denote the least upper bound with  $\sup S$ .
3. We have the same thing for  $\inf S$ .

**Theorem: (Monotone Sequence Theorem)** Every bounded non-decreasing sequence of real numbers converges to a limit.

**Definition: Subsequence** A subsequence of a sequence  $\{\mathbf{a}_j\}_{j \geq j_0} \subseteq \mathbb{R}^n$  is a new sequence denoted  $\{\mathbf{a}_{k_j}\}_j$  where  $\{k_j\}$  is an increasing sequence of integers such that  $k_j \geq j_0, \forall j$ . Intuitively, we can think of this as “following a pattern, we copy down only part of the original sequence to form a new one”.

**Theorem: (Convergence of Subsequence)** If  $\{\mathbf{a}_j\}_j \subseteq \mathbb{R}^n$  is a sequence that is convergent, then any subsequence of  $\{\mathbf{a}_j\}_j$  converges to the same limit. Notice that a catch here is that not any subset qualifies for a subsequence, as stated above, their indices in the original sequence has to follow some pattern, i.e. they themselves have to form a sequence.

## 1.3 Compactness and Applications

**Definition: Compactness** A set  $S \subseteq \mathbb{R}^n$  is said to be compact if every sequence in  $S$  has a subsequence that converges to a limit in  $S$ .

**Theorem: (Bolzano-Weierstrass)** Let  $S$  be a subset of the  $n$  dimensional Euclidean space  $\mathbb{R}^n$ , then  $S$  is compact if and only if  $S$  is closed and bounded. **Remark:** In fact, compactness was a generalization of closed and bounded in spaces other than  $\mathbb{R}^n$ .

**Proposition.** If  $\{\mathbf{x}_j\}_j$  is a convergent sequence in a closed set  $S \subseteq \mathbb{R}^n$ , then the limit of the sequence must belong to  $S$ .

**Theorem: (EVT)** Assume that  $K$  is a compact subset of  $\mathbb{R}^n$ , and that  $f : K \rightarrow \mathbb{R}$  is continuous. Then

the set  $\{f(\mathbf{x}) : \mathbf{x} \in K\}$  is compact

and there exists  $\mathbf{x}^*$  and  $\mathbf{x}_*$  in  $K$  such that

$$f(\mathbf{x}^*) = \sup\{f(\mathbf{x}) : \mathbf{x} \in K\} \wedge f(\mathbf{x}_*) = \inf\{f(\mathbf{x}) : \mathbf{x} \in K\}$$

When we have  $\mathbf{x}^*$  as above, we say that  $f$  attains its supremum, and when there exists  $\mathbf{x}_*$ , we say that  $f$  attains its infimum. This theorem is of crucial importance in circumstances of optimization problems, since we usually have to know that some solution exists in order to have our calculations make sense.

**Definition: Uniform Continuity** A function  $\mathbf{f} : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^k$  is uniformly continuous if

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \left. \begin{array}{l} \mathbf{x}, \mathbf{y} \in S \text{ and} \\ |\mathbf{x} - \mathbf{y}| < \delta \end{array} \right\} \implies |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \epsilon.$$

Notice that this is different from the definition of the normal continuity. In naive words, if a function is continuous, then at every value it can take, there would be some box (whose measures could depend on the value) that captures the function in the sense that the function doesn't escape. However, here we lost the dependence of the input, so there has to be some "box" that works everywhere.

**Checking Uniform Continuity** The epsilon-delta definition, although is rigorous, could be sometimes tedious to check, we develop the following rule: Let  $f$  be a function that is continuous on a closed interval  $[a, b] \subseteq \mathbb{R}$  and differentiable on the open interval  $(a, b)$ . If there exists some number  $M > 0$  such that  $|f'(x)| \leq M, \forall x \in (a, b)$ , then  $f$  is continuous on  $[a, b]$ .

**Theorem: (Uniform CTS w/ Compactness)** If  $K$  is a compact subset of  $\mathbb{R}^n$  and  $f : K \rightarrow \mathbb{R}^k$  is CTS, then  $f$  is uniformly CTS on  $K$ .

## 1.4 The Intermediate Value Theorem

**Definition: Path-Connected** A set  $S \subseteq \mathbb{R}^n$  is path-connected if, for every pair of  $\mathbf{x}, \mathbf{y} \in S$ ,

$$\exists \text{ CTS } \gamma : [0, 1] \rightarrow S, \text{ s.t. } \gamma(0) = \mathbf{x} \wedge \gamma(1) = \mathbf{y}$$

such that  $\gamma(s) \in S, \forall s \in [0, 1]$

**Theorem: (IVT)** Assume that  $S$  is path-connected subset of  $\mathbb{R}^n$  and that  $f : S \rightarrow \mathbb{R}$  is CTS. If  $\mathbf{a}, \mathbf{b}$  are points in  $S$ , and either  $f(\mathbf{a}) < t < f(\mathbf{b})$ , or  $f(\mathbf{b}) < t < f(\mathbf{a})$ , then there exists a point  $\mathbf{c} \in S$ , s.t.  $f(\mathbf{c}) = t$ .

## 2 Differential Calculus

### 2.1 Differentiation of Real-Fcns

**Definition: 1-Var Diff-able** The following are equivalent definitions:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} &= m \\ \iff \lim_{h \rightarrow 0} \frac{f(x+h) - f(x) - mh}{h} &= 0 \\ \iff E(h) := f(x+h) - f(x) - mh \quad \text{satisfies} \quad \lim_{h \rightarrow 0} \frac{E(h)}{h} &= 0 \\ \iff f(x+h) = f(x) + mh + E(h) \quad \text{and} \quad \lim_{h \rightarrow 0} \frac{E(h)}{h} &= 0 \end{aligned}$$

**Definition: Mult-Var Diff-able** Assume that  $f : S \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ , where  $S$  is an open subset. We say that  $f$  is differentiable at  $\mathbf{x} \in S$  if there exists  $\mathbf{m} \in \mathbb{R}^n$  such that

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \mathbf{m} \cdot \mathbf{h}}{|\mathbf{h}|} = 0$$

and when this holds, we define  $\nabla f(\mathbf{x}) = \mathbf{m}$ . **Alternatively**, we say that  $f$  is differentiable at  $\mathbf{x}$  if there exists  $\mathbf{m}$  such that

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{m} \cdot \mathbf{h} + E(\mathbf{h}), \quad \text{where} \quad \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{E(\mathbf{h})}{|\mathbf{h}|} = 0$$

when this holds, we also define  $\nabla f(\mathbf{x}) = \mathbf{m}$ .

**Theorem: (Diff-able Implies CTS)** Assume that  $f : S \rightarrow \mathbb{R}$ , where  $S$  is an open subset of  $\mathbb{R}^n$ , and that  $\mathbf{x} \in S$ . If  $f$  is differentiable at  $\mathbf{x}$ , then  $f$  is continuous at  $\mathbf{x}$ .

**Definition: Partial Derivative** Recall that we have defined  $\mathbf{e}_j$  to be the unit vector in  $\mathbb{R}^n$  in the  $j$ -th coordinate direction. If  $f$  is a function defined on an open subset  $S \subseteq \mathbb{R}^n$ , then at a point  $\mathbf{x} \in S$ , we define

$$\frac{\partial f}{\partial x_j}(\mathbf{x}) := \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h}$$

Computationally, this could be done by treating other variables as constants and proceed as in first year calculus.

**Theorem: (Diff-able Implies Partial Exist)** Let  $f$  be a function  $S \rightarrow \mathbb{R}$ , where  $S$  is an open subset of  $\mathbb{R}^n$ . If  $f$  is differentiable at a point  $\mathbf{x} \in S$ , then  $\frac{\partial f}{\partial x_j}(\mathbf{x})$  exists  $\forall j = 1, \dots, n$ , and in addition,

$$\nabla f(\mathbf{x}) = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) (\mathbf{x})$$

However, do notice that the converse is **not** true. It could certainly be the case that all partials exist at some point at where the function itself is dis-continuous.

**Theorem: (CTS Partial Implies Diff-able)** Assume  $f$  is a function  $S \rightarrow \mathbb{R}$  for some open  $S \subseteq \mathbb{R}^n$ . If all partial derivatives of  $f$  exist and are CTS at every point of  $S$ , then  $f$  is differentiable at every point of  $S$ .

**Definition: Diff Class** A function  $f : S \rightarrow \mathbb{R}$  is said to be of class  $\mathcal{C}^1$  if all partial derivatives of  $f$  exist and are continuous at every point of  $S$ . (Notice that this is stronger than  $\mathcal{C}^0$ , which is just being continuous itself.)

**Theorem: (Directional Derivatives)** If  $f$  is differentiable at a point  $\mathbf{x}$ , then  $\partial_{\mathbf{u}}f(\mathbf{x})$  exists for every *unit* vector  $\mathbf{u} \in \mathbb{R}^n$ , and moreover

$$\partial_{\mathbf{u}}f(\mathbf{x}) = \mathbf{u} \cdot \nabla f(\mathbf{x})$$

**Theorem: (Fastest Increase, Fundamental Principle)** Assume that  $S$  is an open subset of  $\mathbb{R}^n$  and that  $f : S \rightarrow \mathbb{R}$  is differentiable. At any point  $\mathbf{x} \in S$ , the direction of most rapid increase  $\mathbf{u}^*$  is the vector such that

$$\partial_{\mathbf{u}^*} f(\mathbf{x}) = \max\{\partial_{\mathbf{u}} f(\mathbf{x}) : \mathbf{u} \text{ is some unit vec}\}$$

To save ourselves the time to check these conditions we develop the following formulas:

1. If  $\nabla f(\mathbf{x}) = 0$ , then  $\partial_{\mathbf{u}} f(\mathbf{x}) = 0, \forall \mathbf{u}$ , so every unit vector maximizes and minimizes  $\partial_{\mathbf{u}} f(\mathbf{x})$
2. If  $\nabla f(\mathbf{x}) \neq 0$ , then the directional derivative is maximized at

$$\mathbf{u}^* = \frac{\nabla f(\mathbf{x})}{|\nabla f(\mathbf{x})|}$$

## 2.2 Differentiation of Vec-Fcns

**Definition: Jacobian Derivative** Assume that  $S$  is an open subset of  $\mathbb{R}^n$ . Given a function  $\mathbf{f} : S \rightarrow \mathbb{R}^m$  we say that  $\mathbf{f}$  is differentiable at a point  $\mathbf{a} \in S$  if there exists a  $m \times n$  matrix  $M$  such that the familiar definition holds:

$$\mathbf{f}(\mathbf{a} + \mathbf{h}) = \mathbf{f}(\mathbf{a}) + M\mathbf{h} + \mathbf{E}(\mathbf{h}), \quad \text{where } \lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{\mathbf{E}(\mathbf{h})}{|\mathbf{h}|} = \mathbf{0} \in \mathbb{R}^m$$

when this holds, we say that  $M$  is the derivative of  $\mathbf{f}$  at  $\mathbf{a}$ , and we write  $M = D\mathbf{f}(\mathbf{a})$ .

**Theorem: (Jacobian Matrix)** Suppose that  $S$  is an open subset of  $\mathbb{R}^n$ . Then a function  $\mathbf{f} : S \rightarrow \mathbb{R}^m$  is differentiable at a point  $\mathbf{a} \in S$  if and only if the component functions  $f_j$  are differentiable at  $\mathbf{a}$  for every  $j = 1, \dots, m$ . Moreover, the Jacobian Matrix could be calculated by

$$D\mathbf{f}(\mathbf{a}) = \begin{pmatrix} \partial_1 f_1 & \cdots & \partial_n f_1 \\ \partial_1 f_2 & \cdots & \partial_n f_2 \\ \vdots & \ddots & \vdots \\ \partial_1 f_m & \cdots & \partial_n f_m \end{pmatrix}$$

Furthermore, if all partial derivatives  $\partial_i f_j (i = 1, \dots, n \wedge j = 1, \dots, m)$  exist and are CTS in  $S$ , then  $\mathbf{f}$  is differentiable in  $S$ .



**Definition: Differential** Given a differentiable function  $f : S \rightarrow \mathbb{R}$ , where  $S$  is an open subset of  $\mathbb{R}^n$ , at a point  $\mathbf{a} \in S$  we define the differential linear map  $\mathbb{R}^n \rightarrow \mathbb{R}$  by

$$df|_{\mathbf{a}}(\mathbf{h}) = \nabla f(\mathbf{a}) \cdot \mathbf{h}$$

The most common notation for differential would be

$$df = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_n} dx_n$$

**Linear approximation using the differentials:** The definition of the differential implies that if  $f$  is differentiable at  $\mathbf{a}$ , then

$$f(\mathbf{a} + \mathbf{h}) \sim f(\mathbf{a}) + df|_{\mathbf{a}}(\mathbf{h}) \text{ for } \mathbf{h} \text{ small}$$

## 2.3 The Chain Rule

**Theorem: (Chain Rule)** Assume that  $S$  and  $T$  are open subsets of  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , and that we are giving functions  $\mathbf{g} : S \rightarrow \mathbb{R}^m$  and  $\mathbf{f} : T \rightarrow \mathbb{R}^l$ . Assume also that  $\mathbf{a} \in S$  is a point such that  $\mathbf{g}(\mathbf{a}) \in T$ ; thus  $\mathbf{f} \circ \mathbf{g}(\mathbf{x})$  is well-defined for  $\mathbf{x}$  close to  $\mathbf{a}$ . If  $\mathbf{g}$  is differentiable at  $\mathbf{a}$  and  $\mathbf{f}$  is differentiable at  $\mathbf{g}(\mathbf{a})$ , the composite function  $\mathbf{f} \circ \mathbf{g}$  is differentiable at  $\mathbf{a}$ , and

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{a}) = D\mathbf{f}(\mathbf{g}(\mathbf{a})) D\mathbf{g}(\mathbf{a})$$

Notice that although there are infinitely many cases that some people like to remember, using matrix multiplication will always work. So we will save ourselves that time.

**Tangent Plane to Level Set** Suppose  $S$  is an open subset of  $\mathbb{R}^3$  and that  $f : S \rightarrow \mathbb{R}$  is a function that is differentiable at a point  $\mathbf{a} \in S$ . Assume also that  $\nabla f(\mathbf{a}) \neq 0$ . Consider  $C$  to be the level set of  $f$  that passes through  $\mathbf{a}$ . Define

$$\text{The tangent plane to } C \text{ at } \mathbf{a} := \{\mathbf{x} \in \mathbb{R}^3 : (\mathbf{x} - \mathbf{a}) \cdot \nabla f(\mathbf{a}) = 0\}$$

This may seem a bit abstract to apply, so we shall see this in action through an example. Question: “Find the tangent plane to the surface  $C = \{\mathbf{x} \in \mathbb{R}^3 : x^2 - 2xy + 4yz - z^2 = 2\}$ , @  $\mathbf{a} = (1, 1, 1) \in \mathbb{R}^3$ ”. Notice that  $C$  is the level set of  $F(x, y, z) = x^2 - 2xy + 4z - z^2$  at 2. So applying the formula above yields us that the tangent plane is defined by  $\{(x, y, z) \in \mathbb{R}^3 : y + z = 2\}$

## 2.4 The Mean Value Theorem

**Theorem: (MVT)** Assume that  $f$  is real-valued function of class  $\mathcal{C}^1$  defined on an open set  $S \subseteq \mathbb{R}^n$ . For two points  $\mathbf{a}, \mathbf{b} \in S$ , let  $L_{\mathbf{a}, \mathbf{b}}$  denote the line segment that connects them. if  $L_{\mathbf{a}, \mathbf{b}} \subseteq S$ , then there exists  $\mathbf{c} \in L_{\mathbf{a}, \mathbf{b}}$  such that

$$f(\mathbf{b}) - f(\mathbf{a}) = (\mathbf{b} - \mathbf{a}) \cdot \nabla f(\mathbf{c})$$

**Definition: Convex** A set  $S \subseteq \mathbb{R}^n$  is said to be convex if  $\forall \mathbf{a}, \mathbf{b} \in S, L_{\mathbf{a}, \mathbf{b}} \subseteq S$ . That is, formally

$$\forall \mathbf{a}, \mathbf{b} \in S, \forall s \in [0, 1], s\mathbf{b} + (1 - s)\mathbf{a} \in S$$

We like sets with this property in this context since we want the geometric assumption of MVT to be satisfied. Indeed, convex-ness guarantees “straight line connected-ness”.

**Some Important Results:**

1. The intersection of convex sets is convex, while the union of convex sets need not be convex.
2. If a set in  $\mathbb{R}^n$  qualifies for a subspace, then it is convex. In particular, the Kernel and Image are convex.
3. If  $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a function of the form  $L(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ , where  $A$  is an  $m \times n$  matrix and  $\mathbf{b} \in \mathbb{R}^m$  and if  $S$  is convex subset of  $\mathbb{R}^n$ , then

$$L(S) := \{L(\mathbf{x}) : \mathbf{x} \in S\} \text{ is convex}$$

this means that the  $Im(L|_S)$  is convex.

**Theorem: (MVT w/ Cauchy)** Assume that  $S$  is an open, convex subset of  $\mathbb{R}^n$  and that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a function that is differentiable in  $S$ , and moreover that there exists  $M \geq 0$  such that  $|\nabla f(\mathbf{x})| \leq M, \forall \mathbf{x} \in S$ . Then for every  $\mathbf{a}, \mathbf{b} \in S$ ,

$$|f(\mathbf{b}) - f(\mathbf{a})| \leq M|\mathbf{b} - \mathbf{a}|$$

notice that this is a standard consequence of Cauchy Inequality with application to the MVT.

**Theorem: (Everywhere Diminishing Gradient Implies Constant Fcn)** Assume that  $S$  is open, convex subset of  $\mathbb{R}^n$  and that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a function that is differentiable in  $S$ . If  $\nabla f(\mathbf{x}) = \mathbf{0}, \forall \mathbf{x} \in S$ , then  $f$  is constant on  $S$ .

**Theorem: (Stronger Version Of Above)** Assume that  $S$  is open, **path-connected** subset of  $\mathbb{R}^n$  and that  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is a function that is differentiable in  $S$ . If  $\nabla f(\mathbf{x}) = \mathbf{0}, \forall \mathbf{x} \in S$ , then  $f$  is constant on  $S$ .

## 2.5 Higher Order Derivatives

**Definition: Fcns Of Class  $\mathcal{C}^k$**  We say that  $f$  is of class  $\mathcal{C}^k$  if all the  $k$ -th order partial derivatives exist and are CTS everywhere in  $S$ , where  $f$  is defined.

**Theorem: (Clairaut, Second Derivative)** Assume that  $S$  is an open subset of  $\mathbb{R}^n$ . If  $f : S \rightarrow \mathbb{R}$  is  $\mathcal{C}^2$ (Important!), then

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}, \forall i, j \in \{1, \dots, n\} \text{ everywhere in } S$$

**Theorem: (Clairaut, General)** Assume that  $S$  is an open subset of  $\mathbb{R}^n$  and that  $f : S \rightarrow \mathbb{R}$  is of class  $\mathcal{C}^k$ . For any integers  $i_1, \dots, i_k$  between 1 and  $n$ , if  $j_1, \dots, j_k$  is a reordering of the  $i$ 's, then

$$\frac{\partial}{\partial x_{i_k}} \cdots \frac{\partial}{\partial x_{i_1}} f = \frac{\partial}{\partial x_{j_k}} \cdots \frac{\partial}{\partial x_{j_1}} f$$

everywhere in  $S$ .

## 2.6 Taylor's Theorem

**Definition: Hessian Matrix** The Hessian Matrix of a real valued function  $f$  at  $\mathbf{a}$ , denoted  $H(\mathbf{a})$ , is the (square) box of second derivatives of  $f$  whose  $(i, j)$ -th entry is  $\partial_i \partial_j f(\mathbf{a})$ .

**Proposition: Quadratic Taylor Expansion** We have

$$P_{\mathbf{a},2}(\mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2}(H(\mathbf{a})\mathbf{h}) \cdot \mathbf{h}$$

where we remember if we wish our result to be in terms of  $\mathbf{x}$  rather than  $\mathbf{h}$ , we substitute in  $\mathbf{h} = \mathbf{x} - \mathbf{a}$ .

**Theorem: Quadratic Taylor's Theorem** Assume that  $S \subseteq \mathbb{R}^n$  is an open set and that  $f : S \rightarrow \mathbb{R}$  is a function of class  $\mathcal{C}^2$  on  $S$ . Then for  $\mathbf{a} \in S$ ,  $\mathbf{h} \in \mathbb{R}^n$  such that the line segment connecting  $\mathbf{a}$  and  $\mathbf{a} + \mathbf{h}$  is contained in  $S$ , there exists  $\theta \in (0, 1)$  such that

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2}(H(\mathbf{a} + \theta\mathbf{h})\mathbf{h}) \cdot \mathbf{h}$$

holds, then as a result

$$\lim_{\mathbf{h} \rightarrow 0} \frac{R_{\mathbf{a},2}(\mathbf{h})}{|\mathbf{h}|^2} = 0 \quad \text{where } R_{\mathbf{a},2}(\mathbf{h}) = f(\mathbf{a} + \mathbf{h}) - P_{\mathbf{a},2}(\mathbf{h})$$

## 2.7 Critical Points

**Procedure: Classify Critical Points** In solving a question of  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  we could use the following “quick check” approach:

1. Calculate the gradient of  $F$ , equating it to zero to find the critical points
2. Calculate the Hessian of  $F$ , find the corresponding matrices for each critical points, where the Hessian is defined as

$$H(f) = \begin{bmatrix} \partial_{xx}f & \partial_{xy}f = \partial_{yx}f \\ \partial_{xy}f = \partial_{yx}f & \partial_{yy}f \end{bmatrix}$$

3. Calculate the determinant of the hessian, and there are the following cases to consider
  - (a)  $\det H < 0$ , then  $\text{sig}(H) = (1, 1)$  and the point is a saddle point
  - (b)  $\det H > 0$ , then

- i.  $\text{tr}(H) < 0 \implies \text{sig}(H) = (2, 0)$  and the point is a local minimum
- ii.  $\text{tr}(H) > 0 \implies \text{sig}(H) = (0, 2)$  and the point is a local maximum
- (c)  $\det H = 0$ , then the test is inconclusive. We have to do this case by starring at it. (This is a degenerate point, defined below)

**(Non)-Degenerateness** If  $f$  is a  $\mathcal{C}^2$  function and  $\mathbf{a}$  is a critical point of  $f$ , we say that a critical point is **degenerate** if  $\det H(\mathbf{a}) = 0$ , and **non-degenerate** otherwise.