Chapter 1. Point-Set Topology and Calculus*

This course will review basic mathematical tools used in microeconomics, macroeconomics and econometrics. Some popular resources on the materials in this course include Rudin (1976), Dixit (1990), Novshek (1993), Simon and Blume (1994), Sundaram (1996), de la Fuente (2000), Chiang and Wainwright (2005), Osborne (2016), the mathematical appendix of Mas-Colell et al. (1995) and the references therein. Books designed especially for econometrics include Davidson (1994) and Dhrymes (2013) among others. Our main reference books are Rudin (1976) for Chapter 1, Simon and Blume (1994) and Sundaram (1996) for Chapter 2-4, and Casella and Berger (2002) for Chapter 5-6.

In this chapter, we will review some basics for point-set topology in the the Euclidean space and single and multivariable calculus. Key concepts in point-set topology include open, compactness, etc. In calculus, we will define limits, continuity and differentiability, and also define the Riemann integral and the related Riemann-Stieltjes integral.

Since we use this chapter only as a prerequisite for the optimization theory in the coming chapters, we will only cover the topics that are necessary for future developments. Throughout the course, I will emphasize understanding of basic concepts and intuition and application of basic theorems rather than rigorous proofs; you can study the materials in the following order: slides in class, the lecture notes, and the references. Our assignment and exam are based only on the materials in the slides.

We collect some popular notations here for future reference. "$\forall$" means "Any", $\exists$ means "Exist", "viz." or "i.e." means "that is", "e.g." means "for example", "iff" means "if and only if", "$\equiv$" (or $\triangleq$ or $:=$) means "defined as", "$\Longrightarrow$" means "implies" and "$\Longleftrightarrow$" means "is equivalent to". $\square$ is used to signal the end of an example, and $\blacksquare$ the end of a proof. Real numbers are written using lower case italics, e.g., x. Vectors are defined as column vectors and represented using lowercase bold, e.g., $\mathbf{x}$. Matrices are represented using uppercase bold, e.g., $\mathbf{X}$. Sets are represented by uppercase italic, e.g., X, and their elements by lower case italic, e.g., x.

1 Sets and Set Operations

This section reviews some basic concepts about set and set operation. We will use them to define the consumption set and production set. We will also use them in the optimizing decisions of consumers and firms.

*Email: pingyu@hku.hk

Definition 1.1 (Set) *A* ***set*** *A is a collection of distinct objects.*

An element x in A is denoted as $x \in A$. An empty set is often denoted as $\emptyset$. Elements in a set are not ordered and must be distinct, so the following three sets are the same: $\{1, 2\} = \{2, 1\} = \{1, 2, 1\}$.

Remark 1.1 *In mathematics, "collection", "class" and "family" all mean "set". An "object" in a set is often called a "point" although it can be a function defined in the following section or any mathematical object.*

The following are elementary set operations:

1. Subset: set A is contained in B.

   - $A \subset B$: set A is contained in B, and $A \neq B$.
   - $A \subseteq B$: set A is contained in B, and A and B may be equal.

   Remark 1.2 *Quite often, $\subset$ means $\subseteq$. To emphasize $A \neq B$, $\subsetneq$ is often used. We will use the convention of $\subseteq$ and $\subset$.*

   Remark 1.3 *To prove $A = B$, we need only to show $A \subseteq B$ and $B \subseteq A$.*

2. Union: $A \cup B = \{x | x \in A \text{ or } x \in B\}$. All points in either A or B.

   Remark 1.4 *"$|$" is read as "such that", and is often used exchangeably with ":" in the literature and this course.*

3. Intersection: $A \cap B = \{x | x \in A \text{ and } x \in B\}$. All points in both A and B.

4. Complement: $A^c = \{x | x \notin A\}$. All points not in A. Here, a total set is implicitly defined.

5. Relative Complement: $B \setminus A = \{x \in B | x \notin A\} = B \cap A^c$: all points that are in B, but not in A.

   Remark 1.5 *$B \setminus A$ is often denoted as $B - A$, but we will the convention $B \setminus A$.*

The following are useful properties of set operations.

1. Commutatitivity: $A \cup B = B \cup A$ and $A \cap B = B \cap A$
2. Associativity: $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$
3. Distributive Laws: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
4. De Morgan's Law: $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.

2 Functions

A *function* (or *transformation* or *mapping*) is about relationships between the elements of two sets.

Definition 2.1 (Function) *A **function** $f : X \longmapsto Y$ is a rule that associates each element of X with a unique element of Y; in other words, for each $x \in X$ there exists a specified element $y \in Y$, denoted as $f(x)$. x is called the **argument** or the (independent) **variable** of f, and $f(x)$ is called the **value** of f at x. X is called the **domain** of f, and Y the **codomain**. The set*

$$G_f = \{(x, y) \,| x \in X, y = f(x)\} \subseteq X \times Y$$

*is called the **graph** of f, where $X \times Y \equiv \{(x, y) \,| x \in X, y \in Y\}$ is the **Cartesian product** of X and Y.*[1] *For $A \subseteq X$, the set*

$$f(A) = \{f(x) \,| x \in A\} \subseteq Y$$

*is called the **image** of A under f, and for $B \subseteq Y$, the set*

$$f^{-1}(B) = \{x | f(x) \in B\} \subseteq X$$

*is called the **inverse image** (or **pre-image**) of B under f. The set $f(X)$ is called the **range** of f.*

Definition 2.2 *If $f(X) = Y$ the mapping is said to be **surjective** (or **onto**), i.e., every element of the codomain is mapped to by at least one element of the domain; otherwise, f is said to be from X **into** Y. If $f(x_1) = f(x_2) \Longrightarrow x_1 = x_2$, then the mapping is said to be **injective** (or **one-to-one**/1-1), i.e., every element of the codomain is mapped to by at most one element of the domain. A mapping is **bijective** (or **one-to-one correspondence**) if it is both onto and one-to-one.*

Remark 2.1 *The term* function *is usually reserved for cases when the codomain is the set of* real *numbers. That is why we term utility functions and production functions. The term* correspondence *is used for a rule connecting elements of X to elements of Y where the latter are not necessarily unique. For example, f^{-1} is a correspondence, but not a function unless f is one-to-one. If $f^{-1} : f(X) \to X$ is a function, then we call it the **inverse function** of f. The "correspondence" in one-to-one correspondence is an exceptional use of correspondence.*

Remark 2.2 *The elements in Y that cannot be achieved by f seem not interesting, so we can kick them out to make f surjective. In this case, range and codomain mean the same thing. This is why some books use them interchangeably.*

Definition 2.3 (Composite Mapping) *Let $f : X \longmapsto Y$ and $g : Y \longmapsto Z$ are two mappings. The **composite function** (or **mapping**) $g \circ f : X \longmapsto Z$ takes each $x \in X$ to the element $g(f(x)) \in Z$.*

[1] Cartesian product is named after the French philosopher René Descartes (1596-1650) - from his Latinized name Cartesius.

Remark 2.3 *For $C \subseteq Z$, $(g \circ f)^{-1}(C) = f^{-1}\left(g^{-1}(C)\right) = f^{-1} \circ g^{-1}(C)$.*

3 Point-Set Topology in the Euclidean Space

To define the Euclidean space, we first define $\mathbb{R}^n$. $\mathbb{R}^n$ is the Cartesian product of $\mathbb{R}$ with itself n times. $\mathbb{R}^1$ is the real line; $\mathbb{R}^2$ is the plane; $\mathbb{R}^3$ is the three-dimensional space. The **Euclidean space** is $\mathbb{R}^n$ with the **Euclidean structure** imposed on. To emphasize its Euclidean nature, some mathematicians denote the n-dimensional Euclidean space by E^n, but we will still use $\mathbb{R}^n$ to denote it with the understanding that the Euclidean structure has been imposed.

3.1 Euclidean Spaces

The Euclidean structure is best described by the standard **inner product** (also known as the **dot product** or **scalar product**) on $\mathbb{R}^n$. The inner product of any two real n-vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n,$$

where x_i and y_i are ith coordinates of vectors $\mathbf{x}$ and $\mathbf{y}$ respectively, and $\mathbf{x} \cdot \mathbf{y}$ is often written as $\mathbf{x}'\mathbf{y}$ with $\mathbf{x}'$ meaning the transpose of $\mathbf{x}$ or $\langle \mathbf{x}, \mathbf{y} \rangle$. The result is always a real number. The inner product of $\mathbf{x}$ with itself is always non-negative. This product allows us to define the "length" of a vector $\mathbf{x}$ through square root:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

This length function satisfies the required properties of a norm and is called the **Euclidean norm** on $\mathbb{R}^n$. Finally, one can use the norm to define a **metric** (or **distance**) on $\mathbb{R}^n$ by

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

This distance function is called the **Euclidean metric**.

Remark 3.1 *A* ***space*** *is a set plus some structure on it. So a set imposed a metric, a norm or an inner product is called a metric space, a normed space or an inner product space, respectively. The relationship between these spaces is as follows*

$$\mathbb{R}^n \subset \textit{inner product space} \subset \textit{normed space} \subset \textit{metric space}.$$

Of course, we can naturally define a ***subspace*** *as a subset of the original space inheriting its structure.*

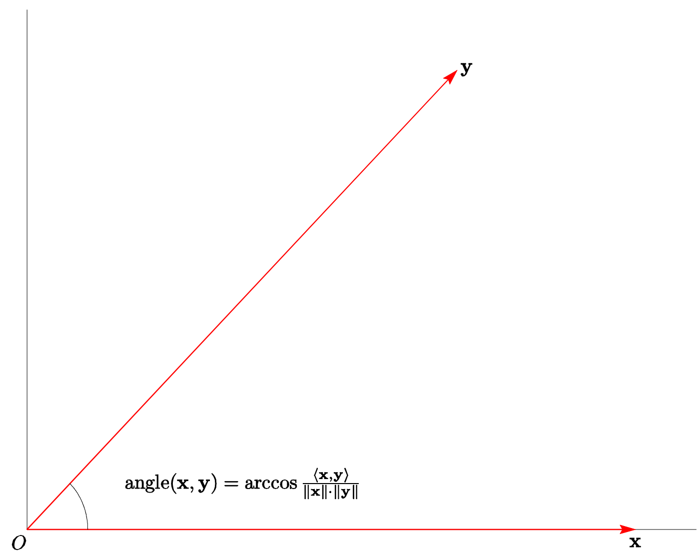

Figure 1: Angle in Two-dimensional Euclidean Space

At this moment, we need to clarify one point - what is the difference between a norm and an inner product? Note that an important inequality in the inner product space is the **Cauchy-Schwarz inequality**:

$$|\langle \mathbf{x}, \mathbf{y}\rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

Due to this inequality, we can define

$$\text{angle}(\mathbf{x}, \mathbf{y}) = \arccos \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|};$$

see Figure 1. We assume the value of the angle is chosen to be in the interval $[0, \pi]$. If $\langle \mathbf{x}, \mathbf{y}\rangle = 0$, angle$(\mathbf{x}, \mathbf{y}) = \frac{\pi}{2}$; we call $\mathbf{x}$ is **orthogonal** to $\mathbf{y}$ and denote it as $\mathbf{x} \perp \mathbf{y}$. Given "orthogonality", we can define "projection". But in a normed space, we cannot always do it. Similarly, in a normed space, we can always define distance, but in a metric space, we cannot always define length.

3.2 Open Sets

Open sets are the basic building blocks of the topological structure of $\mathbb{R}^n$. Roughly speaking, **topology** is about the properties of open sets, and a **topological space** is a set plus a collection of open subsets.[2] To define opens sets, we first define open balls which form the **base** for all open sets in the sense that every open set in $\mathbb{R}^n$ can be written as a union of open balls. An n-dimensional **open ball** (or **open sphere**) of radius r is the collection of points of distance less than r from a

[2] Open sets need not be defined from open balls as below, so a topological space is more general than a metric space.

fixed point in $\mathbb{R}^n$. Explicitly, the open ball with center $\mathbf{x}$ and radius r is defined by

$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n | d(\mathbf{x}, \mathbf{y}) < r\}.$$

The open ball for $n = 1$ is called an **open interval**, and the term **open disk** is sometimes used for $n = 2$ and sometimes as a synonym for open ball.

Definition 3.1 (Open Sets) *A subset U of $\mathbb{R}^n$ is called **open** if for every $\mathbf{x}$ in U there exists an $r > 0$ such that $B_r(\mathbf{x})$ is contained in U. A **neighborhood** of the point $\mathbf{x}$ is any subset N of $\mathbb{R}^n$ that contains an open ball about $\mathbf{x}$ as a subset.*

Remark 3.2 *A set is open iff it contains an open ball around each of its points, i.e., open balls are indeed the base of all open sets.*[3] *Intuitively, an open set is "fat" and does not contain its own "boundary".*

Remark 3.3 *The complement of an open set is called **closed**. Intuitively, a closed set contains its own "boundary".*

Remark 3.4 *Intuitively speaking, a neighborhood of a point is a set of points containing that point where one can move some amount away from that point without leaving the set. Note that the neighborhood N need not be an open set itself. If N is open it is called an **open neighborhood**. Some books require that neighborhoods be open; we will follow this convention.*

Example 3.1 $(0, 1)$ *is open,* $[0, 1]$ *is closed, and* $[0, 1)$ *is neither open nor closed.*

Other important concepts that are releted to open sets include limit points, isolated points, closure, interior, bounary, etc, but we will not disucss them in this course.

3.3 Compactness

We will use a key concept in the first-year courses - compactness. Other concepts such as completeness, separability, connectness, etc, are left for future occasions.

Definition 3.2 (Compactness) *A set E in a metric space (M, d) is called **compact** if each of its open covers has a finite subcover. Otherwise, it is called **non-compact**. Explicitly, this means that for every arbitrary collection*

$$\{U_\alpha\}_{\alpha \in A}$$

of open subsets of M such that

$$E \subseteq \bigcup_{\alpha \in A} U_\alpha,$$

[3] It is not hard to show that if U is open, then $U = \bigcup_{\mathbf{x} \in U} B_{r_\mathbf{x}}(\mathbf{x})$, where we use $r_\mathbf{x}$ to indicate that the radius depends on $\mathbf{x}$.

there is a finite subset J of A such that

$$E \subseteq \bigcup_{i \in J} U_i.$$

*If M is itself compact, (M, d) is said to be a **compact space**.*

Remark 3.5 *A closed subset of a compact space is compact, and a finite union of compact sets is compact.*

Definition 3.3 (Boundedness) *A set E is **bounded** if there is real number r and a point $q \in E$ such that $d(p, q) < r$ for all $p \in E$.*

Remark 3.6 *E is bounded means it can be covered by an open ball $B_r(q)$ of finite radius.*

Theorem 3.1 (Heine-Borel Theorem) *A set $E \subseteq \mathbb{R}^n$ is compact iff it is bounded and closed.*

Example 3.2 *$[0, \infty)$ and $(0, 1)$ are not compact by the Heine-Borel theorem, but $[0, 1]$ is.*

4 Single Variable Calculus

In calculus, we study functions with domain being a subset of $\mathbb{R}^n$. The first-year economic analysis is typically based on "marginal effect analysis", which is usually captured by the derivative of a particular function (e.g., marginal utility, marginal cost, marginal revenue, etc.). This is why we would review properties of continuous and "smooth" functions. The foundation and starting point of calculus is the concept "limit".

4.1 Limits

We first define the limit of a sequence and then the limit of a function.

Definition 4.1 (Sequence and Subsequence) *A **sequence** $\{x_n\}_{n=1}^{\infty}$ is a mapping from $\mathbb{N}$, the set of natural numbers, to some range space. Given a sequence $\{x_n\}$, a **subsequence** of $\{x_n\}$ is defined as $\{x_{n_i}\}_{i=1}^{\infty}$, where $n_1 < n_2 < \cdots$.*

Remark 4.1 *A sequence is automatically ordered, but its **terms** need not be distinct (like a set). Typically the range space is $\mathbb{R}$ although can be extended to any other space.*

Definition 4.2 (Convergence and Limit of a Sequence) *A sequence $\{x_n\}_{n=1}^{\infty}$ is said to **converge** if there is a value $x \in \mathbb{R}$ such that $\forall\ \varepsilon > 0$, $\exists\ n_0 \in \mathbb{N}$ which may depend on ε, such that for all $n > n_0$, $|x_n - x| < \varepsilon$. x is called the **limit** of $\{x_n\}_{n=1}^{\infty}$, and we write $x_n \to x$ or $\lim\limits_{n \to \infty} x_n = x$. If $\{x_n\}_{n=1}^{\infty}$ does not converge, it is said to **diverge**.*

Remark 4.2 *Intuitively, $\lim\limits_{n \to \infty} x_n = x$ means that for n large enough, x_n will stay in an arbitrary small neighborhood of x.*

Example 4.1 *Does the sequence $\{x_n\}_{n=1}^{\infty}$ with $x_n = 1 + (-1)^n/n$ converge?*

The limit of a function is similarly defined as the limit of sequence.

Definition 4.3 (Limit of a Function) *Let $f : [a, b] \to \mathbb{R}$. For any $x \in [a, b]$, we claim $\lim\limits_{t \to x} f(t) = y$ if $\forall\ \varepsilon > 0$, $\exists\ \delta > 0$ which may depend on ε, such that $|f(t) - y| < \varepsilon$ for all t with $|t - x| < \delta$.*

Remark 4.3 *$\lim\limits_{t \to x} f(t) = y$ is equivalent to that for all sequences $\{t_n\}$ (with t_n not equal to x for all n) converging to x a the sequence $f(t_n)$ converges to y.*

4.2 Continuity

Definition 4.4 (Continuity) *Let $f : [a, b] \to \mathbb{R}$. For any $x \in [a, b]$, f is said to be* ***continuous*** *at x if $\lim\limits_{t \to x} f(t) = f(x)$. If f is continuous at every point on $[a, b]$, then f is said to be* ***continuous*** *on $[a, b]$ and denote $f \in C\,[a, b]$.*

Remark 4.4 *Some books define $t = x + \Delta$, so $\lim\limits_{t \to x} f(t) = f(x)$ is equivalently written as $\lim\limits_{\Delta \to 0} f(x + \Delta) = f(x)$. From the remark after the definition of the limit of a function, this can be understood as for* any *sequence $\Delta_n \to 0$, $f(x + \Delta_n) \to f(x)$. Such a notation system can be applied to other definitions such as the derivative.*

An important property of a continuous function is the following intermediate value theorem (IVT).

Theorem 4.1 (Intermediate Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous with $f(a) < f(b)$. Then for any value $M \in (f(a), f(b))$, there is a $c \in (a, b)$ such that $f(c) = M$.*

Remark 4.5 *The case with $f(a) > f(b)$ can be similarly stated. Note that c need not be unique; see Figure 2.*

Remark 4.6 *The intuition of the intermediate value theorem is that the graph of a continuous function on a closed interval can be drawn without lifting your pencil from the paper.*

4.3 Differentiability

Definition 4.5 (Differentiability) *Let $f : [a, b] \to \mathbb{R}$. For any $x \in [a, b]$ form the quotient*

$$\phi(t) = \frac{f(t) - f(x)}{t - x} \quad (a < t < b, t \neq x),$$

and define the ***derivative*** *of f at x as*

$$f'(x) = \lim_{t \to x} \phi(t)$$

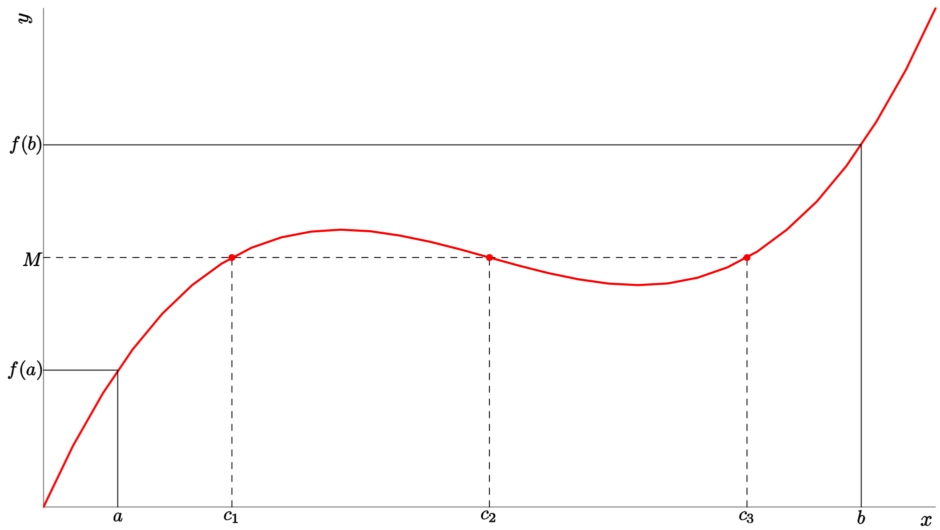

Figure 2: Intermediate Value Theorem

provided this limit exists. If f' is defined at a point x, we say f is ***differentiable*** *at x. If f' is defined at every point of $[a, b]$, we say f is* ***differentiable*** *on $[a, b]$. If f' is further continuous on $[a, b]$, we say f is* ***continuously differentiable*** *or* ***smooth*** *on $[a, b]$ and denote $f \in C^1 [a, b]$.*

Remark 4.7 *From this definition, the derivative is the limit of local slopes.*

Remark 4.8 *The notation of $f'(x)$ is attributed to Newton. The corresponding Leibniz's notation is $\frac{dy}{dx}$ or $\frac{df}{dx}(x)$, where $y \equiv f(x)$. One advantage of Leibniz's notation is that we can intuitively write $dy = f'(x)dx$. If we want to emphasize that the derivative is taken at a specific point, say x_0, then we may write $f'(x_0)$ as $\left.\frac{dy}{dx}\right|_{x=x_0}$.*

Remark 4.9 *A function can be continuous, but not differentiable (easy to find), or differentiable but not continuously differentiable (Exercise).*

Example 4.2 *Suppose $f(x) = x^2$. We want to calculate its derivative at point $x = 2$. Using the definition of derivative, we have*

$$f'(2) = \lim_{t \to 2} \frac{f(t) - f(2)}{t - 2} = \lim_{t \to 2} \frac{t^2 - 2^2}{t - 2} = \lim_{t \to 2} \frac{(t + 2)(t - 2)}{t - 2} = \lim_{t \to 2} (t + 2) = 4.$$

While it is useful to use the definition of derivative to verify the differentiability of a function at a point, it is time-consuming to compute derivatives using the definition. For particular forms of functions, it is easier to remember the formula to compute derivatives. The following table summarizes the derivatives of popular functions.

$f(x)$	$f'(x)$	$f(x)$	$f'(x)$
c	0	$\exp(x)$	$\exp(x)$
cx	c	a^x	$a^x \ln(a)$
x^2	$2x$	$\ln(x)$	$1/x$
x^n	nx^{n-1}	$\log_a(x)$	$1/(x \ln a)$
x^{-1}	$-1/x^2$	$\sin(x)$	$\cos(x)$
$\sqrt{x}$	$\frac{1}{2}\frac{1}{\sqrt{x}}$	$\cos(x)$	$-\sin(x)$

Table: Derivatives of Popular Functions

Theorem 4.2 *Suppose f and g are defined on $[a, b]$ and are differentiable at a point $x \in (a, b)$. Then $f + g$, fg, and f/g are differentiable at x, and*

(i) (Sum Rule) $(f + g)'(x) = f'(x) + g'(x)$;

(ii) (Product Rule) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$;

(iii) (Quotient Rule) $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2}$.

Remark 4.10 *In the quotient rule, if $f = 1$, then we get the* ***reciprocal rule****:* $(1/g)'(x) = -g'(x)/g(x)^2$.

The sign of the derivative of a function can be used to check whether it is monotone.

Definition 4.6 (Monotone Functions) *A function f is said to be* ***non-decreasing*** *(or* ***increasing****) if $f(y) \geq f(x)$ whenever $y > x$. It is* ***non-increasing*** *(or* ***decreasing****) if $-f$ is nondecreasing (increasing). A* ***strictly increasing*** *(****strictly decreasing****) function changes the above inequality to be strict. A* ***monotone*** *(or* ***monotonic****) function is either non-decreasing or non-increasing. A* ***strictly monotone*** *function is either strictly increasing or strictly decreasing.*

Remark 4.11 *Some books use "increasing" for our "strictly increasing".*

Example 4.3 *$f(x) = 2x$ is monotone on $\mathbb{R}$. $f(x) = x^2$ is not monotone on $\mathbb{R}$, but is monotone on $\mathbb{R}_+ \equiv [0, \infty)$.*

Theorem 4.3 *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b).*

(i) *$f'(x) \geq 0$ for $x \in (a, b)$ iff $f(x)$ is non-decreasing;*

(ii) *$f'(x) \leq 0$ for $x \in (a, b)$ iff $f(x)$ is non-increasing;*

(iii) *if $f'(x) > 0$ for $x \in (a, b)$, then $f(x)$ is strictly increasing;*

(iv) *if $f'(x) < 0$ for $x \in (a, b)$, then $f(x)$ is strictly decreasing.*

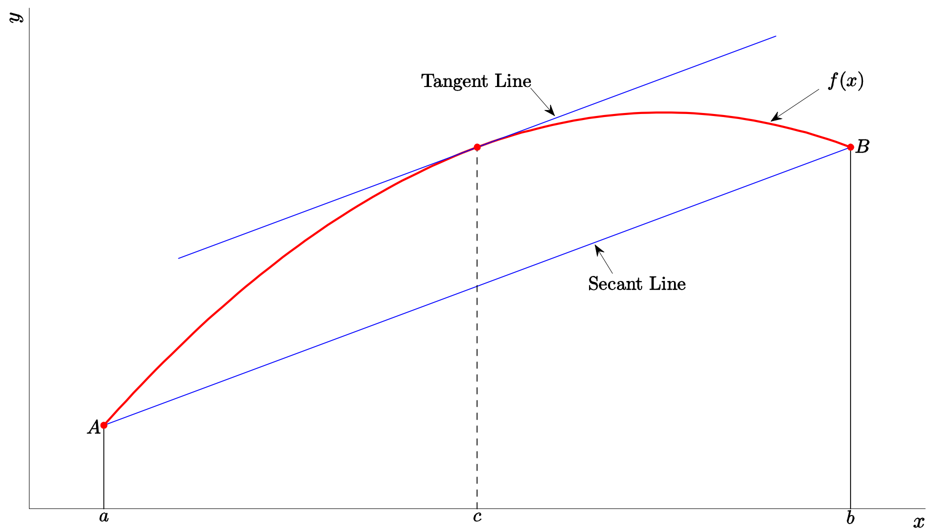

Figure 3: Mean Value Theorem

Remark 4.12 *A strictly increasing function f need not have $f'(x) > 0$ for any $x \in (a, b)$, e.g., for $f(x) = x^3$ on $\mathbb{R}$, $f'(0) = 0$.*

The following theorem, known as the "chain rule" for differentiation. It deals with differentiation of composite functions and is probably the most important theorem about derivatives.

Theorem 4.4 (Chain Rule) *If g is a function that is differentiable at a point c and f is a function that is differentiable at $g(c)$, then the composite function $f \circ g$ is differentiable at c, and the derivative is*

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c),$$

or in short, $(f \circ g)' = (f' \circ g) \cdot g'$.

Another important theorem about differentiation is the mean value theorem (MVT). The mean value theorem states, roughly, that given a planar arc between two endpoints, there is at least one point at which the tangent to the arc is parallel to the secant through its endpoints; see Figure 3.

Theorem 4.5 (Mean Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists some $c \in (a, b)$ such that*

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

We define derivatives using limits. Actually, we can also find limits using derivatives particularly when the limits involve indeterminate forms.

Theorem 4.6 (L'Hôital's Rule) *Suppose functions f and g are differentiable on an open interval I except possibly at a point $c \in I$. If*

$$\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0 \text{ or } \pm\infty,$$

$g'(x) \neq 0$ for all $x \in I$ and $x \neq c$, and $\lim_{x \to c} \frac{f'(x)}{g'(x)}$ exists, then

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}.$$

Remark 4.13 *The differentiation of the numerator and denominator often simplifies the quotient or converts it to a limit that can be evaluated directly.*

Example 4.4 $\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\cos x}{1} = 1.$

4.4 Higher-order Derivatives

If f has a derivative f' on an interval, and if f' is itself differentiable, we denote the derivative of f' by f'' and call f'' the second derivative of f. Continuing in this manner, we obtain functions

$$f, f', f'', f^{(3)}, \cdots, f^{(k)},$$

each of which is the derivative of the preceding one. $f^{(k)}$ is called the k**th derivative**, or the **derivative of order** k, of f. In Leibniz's notation, $f^{(k)}(x) = \frac{d^k y}{dx^k}$, where $y \equiv f(x)$. For a curve, f' means its slope and f'' mean its curvature; that is why monotonicity and concavity of a function (which will be defined in Chapter 3) are related to its first and second order derivatives, respectively.

In order for $f^{(k)}(x)$ to exist at a point x, $f^{(k-1)}(t)$ must exist in a neighborhood for x, and $f^{(k-1)}$ must be differentiable at x. Since $f^{(k-1)}$ must exist in a neighborhood of x, $f^{(k-2)}$ must be differentiable in that neighborhood.

Theorem 4.7 (Taylor's Theorem) *Suppose $f : [a,b] \to \mathbb{R}$, $k \in \mathbb{N}$, $f^{(k-1)} \in C[a,b]$, $f^{(k)}(t)$ exists for any $t \in (a,b)$. Let α, β be distinct points of $[a,b]$, and define*

$$P_{k-1}(t) = \sum_{j=0}^{k-1} \frac{f^{(j)}(\alpha)}{j!}(t-\alpha)^j.$$

Then there exists $x \in (\alpha, \beta)$

$$f(\beta) = P_{k-1}(\beta) + \frac{f^{(k)}(x)}{k!}(\beta - \alpha)^k.$$

Remark 4.14 *Taylor's theorem gives an approximation of a k times differentiable function around a given point by a k-th order Taylor polynomial. The form of the remainder term is attributed* *to Lagrange and is called the* ***Lagrange form*** *(or the mean-value form). Another form of the remainder term takes the* ***integral form****:*

$$R_k(\beta) = \int_\alpha^\beta \frac{f^{(k)}(t)}{(k-1)!}(\beta - t)^{k-1} dt.$$

Yet another form is the ***Peano form*** *of the remainder which writes* $\frac{f^{(k)}(x)}{k!}(\beta-\alpha)^k$ *as* $\left[\frac{f^{(k)}(\alpha)}{k!} + h_k(\beta)\right](\beta-\alpha)^k$*, where* $h_k(\beta)$ *satisfies* $\lim_{\beta\to\alpha} h_k(\beta) = 0$.

Remark 4.15 *Note that Taylor's theorem does not require* β *close to* α *although we typically use it in such a way. Also note that when* $k = 1$*, Taylor's theorem reduces to the mean value theorem.*

4.5 Integrability

Intuitively, the integral of a function is the area "under" the curve defined by the function. Let $f : [a, b] \to \mathbb{R}$. A **partition** of $[a, b]$ is a finite sequence $P = \{x_j\}_{j=0}^n$ such that $a = x_0 < x_1 < \cdots < x_n = b$. Each $[x_i, x_{i+1}]$ is called a **subinterval** of the partition. The **mesh** or **norm** of a partition is defined to be the length of the longest subinterval, that is,

$$\max\{x_i - x_{i-1} | i = 1, \cdots, n\}.$$

A **tagged partition** $P(x, t)$ of an interval $[a, b]$ is a partition together with a finite sequence of numbers $t_0, \cdots, t_{n-1}$ subject to the conditions that for each i, $t_i \in [x_i, x_{i+1}]$. In other words, it is a partition together with a distinguished point of every subinterval. The mesh of a tagged partition is the same as that of an ordinary partition. The **Riemann sum** of f with respect to the tagged partition $x_0, \ldots, x_n$ together with $t_0, \ldots, t_{n-1}$ is

$$\sum_{i=0}^{n-1} f(t_i)(x_{i+1} - x_i).$$

Each term in the sum is the product of the value of the function at a given point and the length of an interval. Consequently, each term represents the (signed) area of a rectangle with height $f(t_i)$ and width $x_{i+1} - x_i$. The Riemann sum is the (signed) area of all the rectangles. The **Riemann integral** is the limit of the Riemann sums of a function as the partitions get finer, and is often denoted as $\int_a^b f(x)dx$. If the limit exists then the function is said to be **Riemann integrable**. The Riemann sum can be made as close as desired to the Riemann integral by making the partition fine enough.

Definition 4.7 (Antiderivative or Indefinite Integral) ***Antiderivative*** *or* ***indefinite integral*** *of a function* f *is a differentiable function* F *whose derivative is equal to the original function* f*. This can be stated symbolically as* $F' = f$.

The fundamental theorem of calculus is a theorem that links the concept of the derivative of a function with the concept of the function's integral. Briefly, differentiation and integration

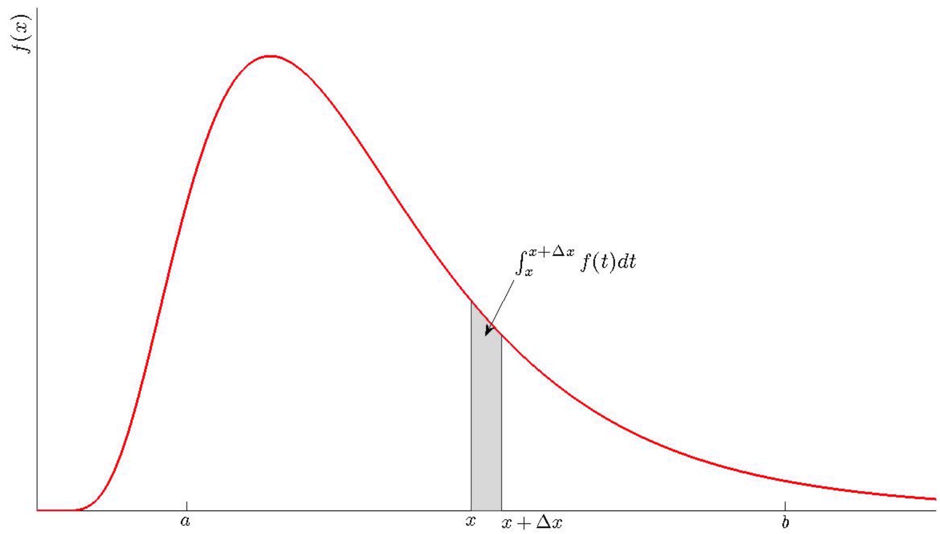

Figure 4: Fundamental Theorem of Calculus: Part I

are inverse operations. There are two parts to the theorem. Loosely put, the first part deals with the derivative of an antiderivative, while the second part deals with the relationship between antiderivatives and definite integrals.

Theorem 4.8 (Fundamental Theorem of Calculus) *Part I: Let* $f : [a, b] \to \mathbb{R}$ *be continuous, and* $F : [a, b] \to \mathbb{R}$ *be defined, for all* $x \in [a, b]$*, by*

$$F(x) = \int_a^x f(t)dt.$$

Then, F *is uniformly continuous on* $[a, b]$*, differentiable on the open interval* (a, b)*, and*

$$F'(x) = f(x)$$

for all $x \in (a, b)$*.*

Part II (Newton–Leibniz Axiom): Let $f : [a, b] \to \mathbb{R}$ *be Riemann integrable, and* $F : [a, b] \to \mathbb{R}$ *be continuous and* $F'(x) = f(x)$ *for all* $x \in (a, b)$*. Then*

$$\int_a^b f(t)dt = F(b) - F(a).$$

Remark 4.16 *Compared with Part I, Part II does not assume that* f *is continuous. Actually, if we assume* f *is continuous, Part II is a straightforward corollary of Part I. Part II is often employed to compute the definite integral of a function* f *for which an antiderivative* F *is known.*

Remark 4.17 *When an antiderivative F exists, then there are infinitely many antiderivatives for f, obtained by adding an arbitrary constant to F. Also, by Part I of the theorem, antiderivatives of f always exist when f is continuous.*

Remark 4.18 *The first part of the theorem, sometimes called the first fundamental theorem of calculus, is that the definite integration of a function is related to its antiderivative, and can be reversed by differentiation. This part of the theorem is important also because it guarantees the existence of antiderivatives for continuous functions. Its intuition is illustrated in Figure 4. The second part of the theorem, sometimes called the second fundamental theorem of calculus, is that the definite integral of a function can be computed by using any one of its infinitely-many antiderivatives. This part of the theorem has key practical applications because it markedly simplifies the computation of definite integrals.*

Integration by parts or **partial integration** relates the integral of a product of functions to the integral of their derivative and antiderivative. It is frequently used to transform the antiderivative of a product of functions into an antiderivative for which a solution can be more easily found.

Theorem 4.9 (Integration by Parts) *Suppose both F and $G : [a, b] \to \mathbb{R}$ are differentiable functions, $F' = f$, and $G' = g$. Then*

$$\int_a^b F(x)g(x)dx = F(x)G(x)|_a^b - \int_a^b f(x)G(x)dx.$$

Some books write $\int_a^b F(x)g(x)dx$ as $\int_a^b F(x)dG(x)$ and $\int_a^b f(x)G(x)dx$ as $\int_a^b G(x)dF(x)$. To understand such notations, we need to define the so-called **Riemann-Stieltjes integral**. The Riemann-Stieltjes integral is closely related to the Riemann integral. The **Riemann-Stieltjes integral** of a function $f : [a, b] \to \mathbb{R}$ with respect to a function $g : [a, b] \to \mathbb{R}$ is denoted by

$$\int_a^b f(x)dg(x)$$

and defined to be the limit, as the norm of the partition

$$P = \{a = x_0 < x_1 < \cdots < x_n = b\}$$

of the interval $[a, b]$ approaches zero, of the approximating sum

$$S(P, f, g) = \sum_{i=0}^{n-1} f(t_i)\left(g(x_{i+1}) - g(x_i)\right),$$

where $t_i \in [x_i, x_{i+1}]$. The two functions f and g are respectively called the **integrand** and the **integrator**. When $g(x) = x$, this is exactly the Riemann integral. We will use the Riemann-Stieltjes integral in Chapter 5 to define the expectation of a random variable.

5 Multivariable Calculus

When $\mathbf{f} : E \to \mathbb{R}^m$, where E is an open set in $\mathbb{R}^n$, we can similarly define whether f is differentiable at a point $\mathbf{x} \in E$. To understand the derivative of such a general function, note that the derivative $f'(x)$ in Definition 4.5 can be equivalently reexpressed as follows: there exists a linear function $f'(x)h$ such that

$$\lim_{h\to 0} \frac{r(h)}{h} \equiv \lim_{h\to 0} \frac{f(x+h) - f(x) - f'(x)h}{h} = 0,$$

where r is for "remainder".

Definition 5.1 (Differentiability) *A function $\mathbf{f} : E \to \mathbb{R}^m$ is said to be* ***differentiable*** *at $\mathbf{x}$ if there exists a linear map $\mathbf{J} : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\lim_{\mathbf{h}\to \mathbf{0}} \frac{\|\mathbf{r}(\mathbf{h})\|_{\mathbb{R}^m}}{\|\mathbf{h}\|_{\mathbb{R}^n}} \equiv \lim_{\mathbf{h}\to \mathbf{0}} \frac{\|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - \mathbf{J}\left(\mathbf{h}\right)\|}{\|\mathbf{h}\|} = 0,$$

where $\mathbf{J}\left(\mathbf{h}\right) = \mathbf{J}\mathbf{h}$, and the $m\times n$ matrix $\mathbf{J}$ is called the ***Jacobian matrix*** *at $\mathbf{x}$. We write $\mathbf{f}'(\mathbf{x}) = \mathbf{J}$. If $\mathbf{f}$ is differentiable at every $\mathbf{x} \in E$, we say f is* ***differentiable*** *in E.*

Remark 5.1 *More often, we use $D_{\mathbf{x}}\mathbf{f}\left(\mathbf{x}\right)$ or $D\mathbf{f}\left(\mathbf{x}\right)$ to denote $\mathbf{f}'(\mathbf{x})$. Note that $\mathbf{J}$ depends on $\mathbf{x}$ as all these notations indicate.*

Remark 5.2 *When $m = n$, $\mathbf{J}$ is square. Both the matrix and its determinant are referred to as the* ***Jacobian*** *in literature.*

Remark 5.3 *The chain rule still holds: If $\mathbf{f} : E \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0 \in E$, $\mathbf{g}$ maps an open set containing $\mathbf{f}(E)$ into $\mathbb{R}^k$, and g is differentiable at $\mathbf{f}(\mathbf{x}_0)$. Then the mapping $\mathbf{F}$ of E into $\mathbb{R}^k$ defined by*

$$\mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$$

is differentiable at $\mathbf{x}_0$, and

$$\mathbf{F}'(\mathbf{x}_0) = \mathbf{g}'(\mathbf{f}(\mathbf{x}_0))\mathbf{f}'(\mathbf{x}_0).$$

The derivative defined above is often called the **total derivative** of $\mathbf{f}$ at $\mathbf{x}$. It may be intriguing how to find $\mathbf{J}$ in practice. It turns out that it can be calculated through partial derivatives.

Definition 5.2 (Partial Derivatives) *Let $\mathbf{f} = (f_1, \cdots, f_m)'$. The* ***partial derivative*** *of f_i at $\mathbf{x}$ with respect to the j-th variable is defined as*

$$\frac{\partial f_i}{\partial x_j}\left(\mathbf{x}\right) = \lim_{h\to 0} \frac{f_i(x_1, \cdots, x_{j-1}, x_j + h, x_{j+1}, \cdots, x_n) - f_i(x_1, \cdots, x_{j-1}, x_j, x_{j+1}, \cdots, x_n)}{h}$$

provided the limit exists.

Remark 5.4 *$\frac{\partial f_i}{\partial x_j}\left(\mathbf{x}\right)$ measures how much f_i would change when all other variables except x_j are fixed at $\mathbf{x}$ and only x_j changes a little bit, so it is very useful in economics for "ceteris paribus" analysis.*

Remark 5.5 *Even if all partial derivatives* $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ *exist at a given point* $\mathbf{x}$*,* $\mathbf{f}$ *need not be (totally) differentiable, or even continuous in the sense that* $\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{f}(\mathbf{x}+\mathbf{h}) = \mathbf{f}(\mathbf{x})$ *(Exercise). On the contrary, if* $\mathbf{f}$ *is differentiable at* $\mathbf{x}$*, then all the partial derivatives at* $\mathbf{x}$ *exist, and*

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix} \equiv \frac{\partial \mathbf{f}}{\partial \mathbf{x}'}(\mathbf{x}),$$

where $\mathbf{J}$ *depends on* $\mathbf{x}$ *in general.*

Remark 5.6 *When* $m = 1$*,* $Df(\mathbf{x})$ *is a* $1 \times n$ *(row) vector; we may intuitively express the total derivative in the form of* ***total differential****,*

$$dy = \frac{\partial f}{\partial x_1}(\mathbf{x})\, dx_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x})\, dx_n,$$

or equivalently,

$$\frac{dy}{dx_1} = \frac{\partial f}{\partial x_1}(\mathbf{x}) + \frac{\partial f}{\partial x_2}(\mathbf{x}) \frac{dx_2}{dx_1} + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}) \frac{dx_n}{dx_1},$$

where $\frac{\partial f}{\partial x_j}(\mathbf{x})$ *is often denoted as* $f_{x_j}(\mathbf{x})$*. Obviously, the total derivative must take into account of the change of* $(x_2, \cdots, x_n)$ *as* x_1 *changes, which is dramatically different from the partial derivative. In other words, the path of* $\mathbf{h} \to \mathbf{0}$ *in* $\mathbb{R}^n$ *in the definition of total derivative is not restricted while the path of* $\mathbf{h} \to \mathbf{0}$ *in the definition of partial derivative* $\partial f_i / \partial x_j(\mathbf{x})$ *is restricted to be along the axis (i.e.,* $h_j \to 0$ *and* $h_k = 0$ *if* $k \neq j$*); see Figure 5 for an intuitive illustration of different pathes of* $\mathbf{h} \to \mathbf{0}$ *in* $\mathbb{R}^2$ *for* $m = 1$*.*

Remark 5.7 *If all partial derivatives exist in a neighborhood of* $\mathbf{x}$ *and are continuous there, then* $\mathbf{f}$ *is (totally) differentiable in that neighborhood and the total derivative is continuous. In this case, it is said that* $\mathbf{f}$ *is a* C^1 *function.*

In the definition of partial derivative, $\mathbf{x} + \mathbf{h}$ converges to $\mathbf{x}$ along the jth coordinate direction. In general, we can check the change of f along any direction.

Definition 5.3 (Directional Derivative) *Let* $f : E \to \mathbb{R}$*, where* E *is an open set in* $\mathbb{R}^n$*. The* ***directional derivative*** *of* f *along a vector* $\mathbf{v} = (v_1, \cdots, v_n)'$ *is the function defined by the limit*

$$\nabla_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}.$$

Remark 5.8 *If the function* f *is differentiable at* $\mathbf{x}$*, then the directional derivative exists along any vector* $\mathbf{v}$*, and one has*

$$\nabla_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}$$

where $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{x})\right)'$ *on the right is called the* ***gradient*** *and* $\cdot$ *is the dot product. Note that* $\nabla f(\mathbf{x}) = Df(\mathbf{x})'$*. Usually,* $\mathbf{v}$ *is normalized such that* $\|\mathbf{v}\| = 1$*.*

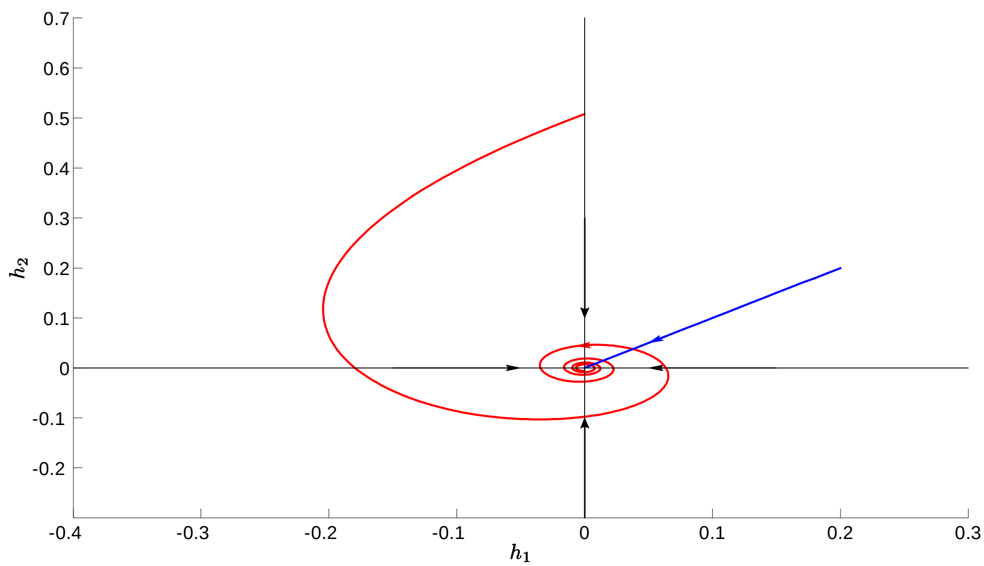

Figure 5: Total Derivative, Partial Derivtive and Directional Derivative: Red Arrow for Total Derivative, Blue Arrow for Directional Derivative, Black Arrows along y-axis for $\partial f/\partial x_2\,(\mathbf{x})$ and Black Arrows along x-axis for $\partial f/\partial x_1\,(\mathbf{x})$

The partial derivative $\partial f_i/\partial x_j$ can be seen as another function defined on E and can again be partially differentiated. To simplify notations, suppose $m = 1$, i.e., $f : E \to \mathbb{R}$. One may wonder whether the order of differentiation matters, that is, whether $\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) \equiv \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(\mathbf{x})\right)$ equals $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \equiv \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}(\mathbf{x})\right)$. This problem is solved in Young's theorem.

Theorem 5.1 (Young's Theorem) *Let $f : E \to \mathbb{R}$, where E is an open set in $\mathbb{R}^n$. If f has continuous second partial derivatives at $\mathbf{x}$, then*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}).$$

Remark 5.9 *If all mixed second order partial derivatives are continuous at a point (or on a set), f is termed a C^2 function at that point (or on that set). Young's theorem implies that for a C^2 function f at $\mathbf{x}$, the **Hessian matrix** f at $\mathbf{x}$, $\mathbf{H}(\mathbf{x})$, is **symmetric**, where*

$$\mathbf{H}(\mathbf{x}) \equiv \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{pmatrix} = \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}'}(\mathbf{x}),$$

with the (i,j)th element being $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$. The Hessian matrix of f at $\mathbf{x}$ is often denoted as $D^2_{\mathbf{x}} f(\mathbf{x})$

or $D^2 f(\mathbf{x})$.

Remark 5.10 *As in the case of the Jacobian, the term "**Hessian**" unfortunately appears to be used both to refer to this matrix and to the determinant of this matrix.*

We are now ready to state Taylor's theorem for multivariate functions. First, we need to introduce some notations. For $\boldsymbol{\alpha} \in \mathbb{Z}_+^n$ and $\mathbf{x} \in \mathbb{R}^n$ with $\mathbb{Z}_+ \equiv \mathbb{N} \cup \{0\}$, define

$$|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_n, \boldsymbol{\alpha}! = \alpha_1! \cdots \alpha_n!, \mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

If all the k-th order partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$ are continuous at $\mathbf{a} \in \mathbb{R}^n$, then by Young's theorem, one can change the order of mixed derivatives at $\mathbf{a}$, so the notation

$$D^{\boldsymbol{\alpha}} f = \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}, |\boldsymbol{\alpha}| \leq k$$

for the higher order partial derivatives is justified in this situation. The same is true if all the $(k-1)$-th order partial derivatives of f exist in some neighborhood of $\mathbf{a}$ and are differentiable at $\mathbf{a}$. Then we say that f is k **times differentiable at the point a**.

Theorem 5.2 (Taylor's Theorem for Multivariate Functions) *Let $f : E \to \mathbb{R}$ be a k times differentiable function on an open convex set $E \subseteq \mathbb{R}^n$. Then for any two distinct points $\mathbf{a}$ and $\mathbf{x}$ in E,*

$$f(\mathbf{x}) = \sum_{|\boldsymbol{\alpha}| \leq k-1} \frac{D^{\boldsymbol{\alpha}} f(\mathbf{a})}{\boldsymbol{\alpha}!} (\mathbf{x} - \mathbf{a})^{\boldsymbol{\alpha}} + R_k(\mathbf{x}),$$

where

$$R_k(\mathbf{x}) = \sum_{|\boldsymbol{\alpha}| = k} D^{\boldsymbol{\alpha}} f(c\mathbf{a} + (1-c)\mathbf{x}) \frac{(\mathbf{x} - \mathbf{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!}$$

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

Remark 5.11 *$R_k(\mathbf{x})$ takes the Lagrange form. We can also state the remainder term in the integral form as*

$$R_k(\mathbf{x}) = k \sum_{|\boldsymbol{\alpha}| = k} \frac{(\mathbf{x} - \mathbf{a})^{\boldsymbol{\alpha}}}{\boldsymbol{\alpha}!} \int_0^1 (1-t)^{k-1} D^{\boldsymbol{\alpha}} f(\mathbf{a} + t(\mathbf{x} - \mathbf{a}))\, dt,$$

or in the Peano form as $R_k(\mathbf{x}) = \sum_{|\boldsymbol{\alpha}| = k} \left[\frac{D^{\boldsymbol{\alpha}} f(\mathbf{a})}{\boldsymbol{\alpha}!} + h_\alpha(\mathbf{x}) \right] (\mathbf{x} - \mathbf{a})^{\boldsymbol{\alpha}}$ *with* $\lim_{\mathbf{x} \to \mathbf{a}} h_\alpha(\mathbf{x}) = 0$.

Remark 5.12 *When $k = 1$, we get the mean value theorem:*

$$f(\mathbf{x}) = f(\mathbf{a}) + \nabla f(c\mathbf{a} + (1-c)\mathbf{x}) \cdot (\mathbf{x} - \mathbf{a})$$

for some $c \in (0, 1)$. When $k = 2$, we can approximate $f(\mathbf{x})$ by a quadratic function in the neighborhood of $\mathbf{a}$:

$$f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2} (\mathbf{x} - \mathbf{a})' \mathbf{H} (\mathbf{x} - \mathbf{a}),$$

where $\mathbf{H}$ *is evaluated at* $c\mathbf{a} + (1-c)\mathbf{x}$ *for some* $c \in (0,1)$.

We can similarly define the **multiple integral**

$$\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(\mathbf{x}) dx_1 \cdots dx_n$$

as in the single variable case. Now, we need to partition each interval $[a_i, b_i]$, $i = 1, \cdots, n$, to construct subrectangles. In constructing the Riemann sum, replace the width of subintervals by the volume of subrectangles, and let t_i stay in the corresponding subrectangle. Then make the width of all subintervals uniformly converge to zero and check whether the Riemann sum converges.

Integrals of a function of two variables over a region in $\mathbb{R}^2$ are called **double integrals**, and integrals of a function of three variables over a region of $\mathbb{R}^3$ are called **triple integrals**. They can be used to calculate areas and volumes of regions in the plane, respectively.

In single variable calculus, the fundamental theorem of calculus establishes a link between the derivative and the integral. The link between the derivative and the integral in multivariable calculus is embodied by the integral theorems of vector calculus: Gradient theorem, Stokes' theorem, Divergence theorem, and Green's theorem. In a more advanced study of multivariable calculus, it is seen that these four theorems are specific incarnations of a more general theorem, the **generalized Stokes' theorem**. These theorems are rarely used in economics, so are neglected in this course.

We finally state a very useful result that shows how to differentiate under the integral sign.

Theorem 5.3 (Leibniz's Rule) *Let* $f(x,t)$ *be a function such that both* $f(x,t)$ *and its partial derivative* $(\partial/\partial t) f(x,t)$ *are continuous in* t *and* x *in some region of the* (x,t)*-plane, including* $\alpha(t) \leq x \leq \beta(t)$, $t_0 \leq t \leq t_1$. *Also suppose that the functions* $\alpha(t)$ *and* $\beta(t)$ *are both continuous and both have continuous derivatives for* $t_0 \leq t \leq t_1$. *Then for* $t_0 \leq t \leq t_1$,

$$F(t) \equiv \int_{\alpha(t)}^{\beta(t)} f(x,t)\, dx$$

is differentiable and

$$\frac{d}{dt} F(t) = f(\beta(t), t)\, \beta'(t) - f(\alpha(t), t)\, \alpha'(t) + \int_{\alpha(t)}^{\beta(t)} f_t(x,t)\, dx.$$