Lecture 1: The Real Number System

In this note we will give some idea about the real number system and its properties.

We start with the set of integers. We know that given any two integers, these can be added, one can be subtracted from the other and they can be multiplied. The result of each of these operations is again an integer. Further, if p and q are integers and $p - q > 0$, then we say that $p > q$ (this defines an order relation '$<$' on the set of integers.)

Next we consider numbers of the form m/n, where m and n are integers and $n \neq 0$, called rational numbers. (We may assume that m and n have no common factor.) The operations of addition (and subtraction), multiplication and '$<$' extend to this set in a natural way. We shall see, from a very simple situation that numbers other than rational numbers are needed. Consider a square whose side has unit length. Then by Pythagoras Theorem, the length l of the diagonal must satisfy $l^2 = 2$ (we write $l = \sqrt{2}$). What is l? Suppose $l = m/n$, where m and n are integers, which are not both even. Then $l^2n^2 = 2n^2 = m^2$. Thus m^2 is even. Since the square of an odd integer is odd, we conclude that m is even, so that n^2 is even. Hence n is divisible by 2. This contradicts our assumption.

The above discussion shows that $\sqrt{2}$ is not a rational number and we need numbers such as $\sqrt{2}$. Such numbers will be called irrational numbers. The problem now is to define these numbers from $\mathbb{Q}$.

We will not present the mathematical definition of real numbers here, *since it is bit involved.* Instead we will give a rough idea about real numbers.

On a straight line, if we mark off segments ..., $[-1, 0], [0, 1], [1, 2], \ldots$ then all the rational numbers can be represented by points on this straight line. The set of points representing rational numbers seems to fill up this line (rational number $\frac{r+s}{2}$ lies in between the rational numbers r and s). But we have seen above that the rationals do not cover the entire straight line. Intuitively we feel that there should be a larger set of numbers, say $\mathbb{R}$ such that there is a correspondence between $\mathbb{R}$ and the points of this straight line. Indeed, one can construct such a set of numbers from the rational number system $\mathbb{Q}$, called set of real numbers, which contains the set of rationals and also numbers such as $\sqrt{2}, \sqrt{3}, \sqrt{5}$ and more. Moreover, on this set we can define operations of addition and multiplication, and an order in such a way that when these operations are restricted to the set of rationals, they coincide with the usual operations and the usual order. The set $\mathbb{R}$ with these operations is called the real number system.

An important property of $\mathbb{R}$, which is missing in $\mathbb{Q}$ is the following.

Completeness property of real number system:

A subset A of $\mathbb{R}$ is said to be *bounded above* if there is an element $x_0 \in \mathbb{R}$ such that $x \leq x_0$ for all $x \in A$. Such an element x_0 is called an *upper bound* of A . Similarly A is said to be *bounded below* if there exists $y_0 \in \mathbb{R}$ such that $y_0 \leq x$ for all $x \in \mathbb{R}$.

An upper bound x_0 of A is said to be a *least upper bound* (l.u.b.) or supremum (sup) of A if whenever z is an upper bound of A, $x_0 \leq z$. A *greatest lower bound* (g.l.b.) or infimum (inf) is defined similarly.

Remark. The least upper bound or the greatest lower bound may not belong to the set A. For example, 1 is the l.u.b of the sets $\{x : 0 < x < 1\}, \{x : 0 \leq x \leq 1\}$ and $\{1 - \frac{1}{n} : n \in \mathbb{N}\}$.

Real number system has the property that every non-empty subset of $\mathbb{R}$ which is bounded above has a least upper bound. This property is called *least upper bound property.* The greatest lower bound property is defined similarly. By *completeness property* we mean either l.u.b. property or g.l.b. property.

We will now show that the following properties which will be used later are important consequences of the completeness property of $\mathbb{R}$.

Proposition 1.1 (Archimedean property): *If $x, y \in \mathbb{R}$ and $x > 0$, then there is a positive integer n such that $nx > y$.*

Proof (*): Suppose that $nx \leq y$ for every positive integer n. Then y is an upper bound of the set $A = \{nx : n \in \mathbb{N}\}$. By the least upper bound property, let α be a l.u.b. of A. Then $(n+1)x \leq \alpha$ for all n and so $nx \leq \alpha - x < \alpha$ for all n i.e. $\alpha - x$ is also an upper bound which is smaller than α. This is a contradiction. $\square$

Remark : Let $A = \{r \in \mathbb{Q} : r > 0, r^2 < 2\}$; this is a non-empty and bounded subset of $\mathbb{Q}$. The set A does not have l.u.b. in $\mathbb{Q}$ (see Problem 10 in Practice Problems 1). This shows that $\mathbb{Q}$ does not have the least upper bound property.

The Archimedean property leads to the *"density of rationals in $\mathbb{R}$"* and *"density of irrationals in $\mathbb{R}$"*.

Proposition 1.2: *Between any two distinct real numbers there is a rational number.*

Proof : Suppose $x, y \in \mathbb{R}$, $y - x > 0$. We have to find two integers m and n, $n \neq 0$ such that

$$x < \frac{m}{n} < y \quad \text{i.e.,} \quad x < \frac{m}{n} < x + (y - x).$$

Now by the Archimedean property there exists a positive integer n such that $n(y - x) > 1$. Then we can find an integer m lying between nx and $ny = nx + n(y - x)$ (see Problem 8 of Practice Problems 1). This proves the result. $\square$

Problem 1 : *Between any two distinct real numbers there is an irrational number.*

Solution : Suppose $x, y \geq 0$, $y - x > 0$. Then $\frac{1}{\sqrt{2}}x < \frac{1}{\sqrt{2}}y$. By Proposition 1.2, there exists a rational number r such that $x < r\sqrt{2} < y$. $\square$

Problem 2 : *Find the supremum and the infimum of the set $\left\{\frac{m}{m+n} : m, n \in \mathbb{N}\right\}$.*

Solution : First note that $0 < \frac{m}{m+n} < 1$. We guess that inf $= 0$ because $\frac{1}{1+n}$ is in the set and it approaches 0 when n is very large. Formally to show that 0 is the infimum, we have to show that 0 is a lower bound and it is the least among all the lower bounds of the set. It is clear that 0 is a lower bound. It remains to show that a number $\alpha > 0$ cannot be a lower bound of the given set. This is true because we can find an n such that $\frac{1}{1+n} < \alpha$ using the Archimedean property. Note that $\frac{1}{1+n}$ is in the given set! Similarly we can show that sup $= 1$. $\square$

The following problem will be used later.

Problem 3 : *Let A be a nonempty subset of $\mathbb{R}$ and α a real number. If $\alpha = supA$ then $a \leq \alpha$ for all $a \in A$ and for any $\varepsilon > 0$, there is some $a_0 \in A$ such that $\alpha - \varepsilon < a_0$.*

Solution : Suppose $\alpha = \sup A$. Since it is an upper bound we have $a \leq \alpha$ for all $a \in A$. Suppose $\varepsilon > 0$. If there is no $a \in A$ such that $\alpha - \varepsilon < a$, then we have $a \leq \alpha - \varepsilon < \alpha$ for all $a \in A$. This implies that $\alpha - \varepsilon$ is an upper bound. This contradicts the fact that α is the least upper bound. $\square$

Lecture 2 : Convergence of a Sequence, Monotone sequences

In less formal terms, a sequence is a set with an order in the sense that there is a first element, second element and so on. In other words for each positive integer 1,2,3, ..., we associate an element in this set. In the sequel, we will consider only sequences of real numbers.

Let us give the formal definition of a sequence.

Definition : A function $f : \{1, 2, 3, \ldots\} \to \mathbb{R}$ is called a sequence of real numbers. We write $f(n) = x_n$, then the sequence is denoted by $x_1, x_2, \ldots$, or simply by (x_n). We call x_n the nth term of the sequence or the value of the sequence at n.

Some examples of sequences:

1. $(n) = 1, 2, 3, \ldots$
2. $(\frac{1}{n}) = 1, \frac{1}{2}, \frac{1}{3}, \ldots$
3. $(\frac{(-1)^n}{n}) = -1, \frac{1}{2}, \frac{-1}{3}, \ldots$
4. $(1 - \frac{1}{n}) = 0, 1 - \frac{1}{2}, 1 - \frac{1}{3}, \ldots$
5. $(1 + \frac{1}{10^n}) = 1.1, 1.01, 1.001, \ldots$
6. $((-1)^n) = -1, +1, -1, +1, \ldots$

Before giving the formal definition of convergence of a sequence, let us take a look at the behaviour of the sequences in the above examples.

The elements of the sequences $(\frac{1}{n})$, $(1 - 1/n)$ and $(1 + 1/10^n)$ seem to "approach" a single point as n increases. In these sequences the values are either increasing or decreasing as n increases, but they "eventually approach" a single point. Though the elements of the sequence $((-1)^n/n)$ oscillate, they "eventually approach" the single point 0. The common feature of these sequences is that the terms of each sequence "accumulate" at only one point. On the other hand, values of the sequence (n) become larger and larger and do not accumulate anywhere. The elements of the sequence $((-1)^n)$ oscillate between two different points -1 and 1; i.e., the elements of the sequence come close to -1 and 1 "frequently" as n increases.

Convergence of a Sequence

Let us distinguish sequences whose elements approach a single point as n increases (in this case we say that they converge) from those sequences whose elements do not. Geometrically, it is clear that if the elements of the sequence (x_n) come eventually inside *every* ϵ-neighbourhood $(x_0 - \epsilon, x_0 + \epsilon)$ of x_0 then (x_n) approaches x_0.

Let us now state the formal definition of convergence.

Definition : We say that a sequence (x_n) converges if there exists $x_0 \in I\!R$ such that for every $\epsilon > 0$, there exists a positive integer N (depending on ϵ) such that

$$x_n \in (x_0 - \epsilon, x_0 + \epsilon) \quad (\text{or } |x_n - x_0| < \epsilon) \text{ for all } n \geq N.$$

It can be easily verified that if such a number x_0 exists then it is unique. In this case, we say that the sequence (x_n) converges to x_0 and we call x_0 the limit of the sequence (x_n). If x_0 is the limit of (x_n), we write $\lim_{n\to\infty} x_n = x_0$ or $x_n \to x_0$.

Examples : 1. Let us show that the sequence $(\frac{1}{n})$ in Example 1 has limit equal to 0. For arbitrary $\epsilon > 0$, the inequality

$$|x_n| = \frac{1}{n} < \epsilon$$

is true for all $n > \frac{1}{\epsilon}$ and hence for all $n > N$, where N is any natural number such that $N > \frac{1}{\epsilon}$. Thus for any $\epsilon > 0$, there is a natural number N such that $|x_n| < \epsilon$ for every $n \geq N$.

2. The sequence in Example 4 converges to 1, because in this case

$$|1 - x_n| = |1 - \frac{n-1}{n}| = \frac{1}{n} \leq \epsilon$$

for all $n > N$ where N is any natural number greater than $\frac{1}{\epsilon}$.

Remark: The convergence of each sequence given in the above examples is verified directly from the definition. In general, verifying the convergence directly from the definition is a difficult task. We will see some methods to find limits of certain sequences and some sufficient conditions for the convergence of a sequence.

The following three results enable us to evaluate the limits of many sequences.

Limit Theorems

Theorem 2.1: *Suppose $x_n \to x$ and $y_n \to y$. Then*

1. $x_n + y_n \to x + y$
2. $x_n y_n \to xy$
3. $\frac{x_n}{y_n} \to \frac{x}{y}$ if $y \neq 0$ and $y_n \neq 0$ for all n.

Example : Let $x_n = \frac{1}{1^2+1} + \frac{1}{2^2+2} + \cdots + \frac{1}{n^2+n}$. Then $x_n = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \cdots + \frac{1}{n} - \frac{1}{n+1} \to 1$

Theorem 2.2 : (*Sandwich Theorem*) *Suppose that $(x_n), (y_n)$ and (z_n) are sequences such that $x_n \leq y_n \leq z_n$ for all n and that $x_n \to x_0$ and $z_n \to x_0$. Then $y_n \to x_0$.*

Proof: Let $\epsilon > 0$ be given. Since $x_n \to x_0$ and $z_n \to x_0$, there exist N_1 and N_2 such that

$$x_n \in (x_0 - \epsilon, x_0 + \epsilon) \quad \text{for all} \quad n \geq N_1$$

and

$$z_n \in (x_0 - \epsilon, x_0 + \epsilon) \quad \text{for all} \quad n \geq N_2.$$

Choose $N = max\{N_1, N_2\}$. Then, since $x_n \leq y_n \leq z_n$, we have

$$y_n \in (x_0 - \epsilon, x_0 + \epsilon) \quad \text{for all} \quad n \geq N.$$

This proves that $y_n \to x_0$. □

Examples : 1. Since $\frac{-1}{n} \leq \frac{\sin n}{n} \leq \frac{1}{n}$, by sandwich theorem $\frac{\sin n}{n} \to 0$.

2. Let $y_n = \frac{n^2}{n^3+n+1} + \frac{n^2}{n^3+n+2} + \cdots + \frac{n^2}{n^3+2n}$. Then $\frac{n.n^2}{n^3+2n} \leq y_n \leq \frac{n.n^2}{n^3+n+1}$ and hence $y_n \to 1$.

3. Let $x \in \mathbb{R}$ and $0 < x < 1$. We show that $x^n \to 0$. Write $x = \frac{1}{1+a}$ for some $a > 0$. Then by Bernoulli's inequality, $0 < x^n = \frac{1}{(1+a)^n} \leq \frac{1}{1+na} < \frac{1}{na}$. By sandwich theorem $x^n \to 0$.

4. Let $x \in \mathbb{R}$ and $x > 0$. We show that $x^{\frac{1}{n}} \to 1$. Suppose $x > 1$ and $x^{\frac{1}{n}} = 1 + d_n$ for some $d_n > 0$. By Bernoulli's inequality, $x = (1 + d_n)^n > 1 + nd_n > nd_n$ which implies that $0 < d_n < \frac{x}{n}$ for all $n \in \mathbb{N}$. By sandwich theorem $d_n \to 0$ and hence $x^{\frac{1}{n}} \to 1$. If $0 < x < 1$, let $y = \frac{1}{x}$ so that $y^{\frac{1}{n}} \to 1$ and hence $x^{\frac{1}{n}} \to 1$.

5. We show that $n^{\frac{1}{n}} \to 1$. Let $n^{\frac{1}{n}} = 1 + k_n$ for some $k_n > 0$ when $n > 1$. Hence $n = (1 + k_n)^n > 1$ for $n > 1$. By Binomial theorem, if $n > 1$, $n \geq 1 + \frac{1}{2}n(n-1)k_n^2$. Therefore $n - 1 \geq \frac{1}{2}n(n-1)k_n^2$ and hence $k_n^2 \leq \frac{2}{n}$. By sandwich theorem $k_n \to 0$ and therefore $n^{\frac{1}{n}} \to 1$.

The following result, called ratio test for sequences, can be applied to certain type of sequences for convergence.

Theorem 2.3: *Let (x_n) be a sequence of real numbers such that $x_n > 0$ for all n and $\lim\limits_{n\to\infty} \frac{x_{n+1}}{x_n} = \lambda$. Then*

1. *if $\lambda < 1$ then $\lim\limits_{n\to\infty} x_n = 0$,*
2. *if $\lambda > 1$ then $\lim\limits_{n\to\infty} x_n = \infty$.*

Proof : 1. Since $\lambda < 1$, we can find an r such that $\lambda < r < 1$. As $\lim\limits_{n\to\infty} \frac{x_{n+1}}{x_n} = \lambda$, there exists n_0 such that $\frac{x_{n+1}}{x_n} < r$ for all $n \geq n_0$. Hence,

$$0 < x_{n+n_0} < rx_{n+n_0-1} < r^2 x_{n+n_0-2} < \cdots < r^n x_{n_0}.$$

Note that $\lim\limits_{n\to\infty} r^n = 0$ as $0 < r < 1$. So by the sandwich theorem $x_n \to 0$.

2. Since $\lambda > 1$, we can find $r \in \mathbb{R}$, such that $1 < r < \lambda$. Arguing along the same lines as in 1., we get $n_0 \in \mathbb{N}$, such that $x_{n+1} > rx_n$ for all $n \geq n_0$. Similarly, $x_{n+n_0} > r^n x_{n_0}$. Since $r > 1$, $\lim\limits_{n\to\infty} r^n = \infty$ and therefore $\lim\limits_{n\to\infty} x_n = \infty$. $\square$

Examples : 1. Let $x_n = \frac{n}{2^n}$ and $y_n = \frac{2^n}{n!}$. Then $x_n \to 0$ as $\lim\limits_{n\to\infty} \frac{x_{n+1}}{x_n} = \frac{1}{2}$. We do similarly for y_n.

2. Let $x_n = ny^{n-1}$ for some $y \in (0, 1)$. Since $\lim\limits_{n\to\infty} \frac{x_{n+1}}{x_n} = y$, $x_n \to 0$.

3. Let $x_n = \frac{n^s}{(1+p)^n}$ for some $s > 0$ and $p > 0$. Repeat the argument as in the previuos problem and show that $x_n \to 0$.

4. Let $b > 1$ and $x_n = \frac{b^n}{n^2}$. Then $\lim\limits_{n\to\infty} \frac{x_{n+1}}{x_n} = b$. Therefore, $\lim\limits_{n\to\infty} x_n = \infty$.

5. In the previous theorem if $\lambda = 1$ then we cannot make any conclusion. For example, consider the sequences $(n), (\frac{1}{n})$ and $(2 + \frac{1}{n})$.

In the previous results we could guess the limit of a sequence by comparing the given sequence with some other sequences whose limits are known and then we could verify that our guess is correct. We now give a simple criterion for the convergence of a sequence (without having any knowledge of its limit).

Before presenting a criterion (a sufficient condition), let us see a necessary condition for the convergence of a sequence.

Theorem 2.4: *Every convergent sequence is a bounded sequence, that is the set $\{x_n : n \in \mathbb{N}\}$ is bounded.*

Proof : Suppose a sequence (x_n) converges to x. Then, for $\epsilon = 1$, there exist N such that

$$|x_n - x| \leq 1 \quad \text{for all} \;\; n \geq N.$$

This implies $|x_n| \leq |x| + 1$ for all $n \geq N$. If we let

$$M = \max\{|x_1|, |x_2|, \ldots, |x_{N-1}|\},$$

then $|x_n| \leq M + |x| + 1$ for all n. Hence (x_n) is a bounded sequence. □

Remark : The condition given in the previous result is necessary but not sufficient. For example, the sequence $((-1)^n)$ is a bounded sequence but it does not converge.

One naturally asks the following question:

Question : Boundedness $+$ (??) $\Rightarrow$ **Convergence.**

We now find a condition on a bounded sequence which ensures the convergence of the sequence.

Monotone Sequences

Definition : We say that a sequence (x_n) is increasing if $x_n \leq x_{n+1}$ for all n and strictly increasing if $x_n < x_{n+1}$ for all n. Similarly, we define decreasing and strictly decreasing sequences. Sequences which are either increasing or decreasing are called monotone.

The following result is an application of the least upper bound property of the real number system.

Theorem 2.5: *Suppose (x_n) is a bounded and increasing sequence. Then the least upper bound of the set $\{x_n : n \in \mathbb{N}\}$ is the limit of (x_n).*

Proof: Suppose $\sup\limits_n x_n = M$. Then for given $\epsilon > 0$, there exists n_0 such that $M - \epsilon \leq x_{n_0}$. Since (x_n) is increasing, we have $x_{n_0} \leq x_n$ for all $n \geq n_0$. This implies that

$$M - \epsilon \leq x_n \leq M \leq M + \epsilon \;\; \text{for all } n \geq n_0.$$

That is $x_n \to M$. □

For decreasing sequences we have the following result and its proof is similar.

Theorem 2.6: *Suppose (x_n) is a bounded and decreasing sequence. Then the greatest lower bound of the set $\{x_n : n \in \mathbb{N}\}$ is the limit of (x_n).*

Examples: 1. Let $x_1 = \sqrt{2}$ and $x_n = \sqrt{2 + x_{n-1}}$ for $n > 1$. Then use induction to see that $0 \leq x_n \leq 2$ and (x_n) is increasing. Therefore, by previous result (x_n) converges. Suppose $x_n \to \lambda$. Then $\lambda = \sqrt{2 + \lambda}$. This implies that $\lambda = 2$.

2. Let $x_1 = 8$ and $x_{n+1} = \frac{1}{2}x_n + 2$. Note that $\frac{x_{n+1}}{x_n} < 1$. Hence the sequence is decreasing. Since $x_n > 0$, the sequence is bounded below. Therefore (x_n) converges. Suppose $x_n \to \lambda$. Then $\lambda = \frac{\lambda}{2} + 2$. Therefore, $\lambda = 2$.

Lecture 3 : Cauchy Criterion, Bolzano-Weierstrass Theorem

We have seen one criterion, called monotone criterion, for proving that a sequence converges without knowing its limit. We will now present another criterion.

Suppose that a sequence (x_n) converges to x. Then for $\epsilon > 0$, there exists an N such that $|x_n - x| < \epsilon/2$ for all $n \geq N$. Hence for $n, m \geq N$ we have

$$|x_n - x_m| = |x_n - x + x - x_m| \leq |x_n - x| + |x - x_m| < \epsilon.$$

Thus we arrive at the following conclusion:

If a sequence (x_n) converges then it satisfies the **Cauchy's criterion:** *for $\epsilon > 0$, there exists N such that $|x_n - x_m| < \epsilon$ for all $n, m \geq N$.*

If a sequence converges then the elements of the sequence get close to the limit as n increases. In case of a sequence satisfying Cauchy criterion the elements get close to each other as m, n increases.

We note that a sequence satisfying Cauchy criterion is a bounded sequence (verify!) with some additional property. Moreover, intuitively it seems as if it converges. We will show that a sequence satisfying Cauchy criterion does converge. We need some results to prove this.

Theorem 3.1 : ***(Nested interval Theorem)*** *For each n, let $I_n = [a_n, b_n]$ be a (nonempty) bounded interval of real numbers such that*

$$I_1 \supset I_2 \supset \cdots \supset I_n \supset I_{n+1} \supset \cdots$$

and $\lim\limits_{n\to\infty}(b_n - a_n) = 0$. Then $\bigcap\limits_{n=1}^{\infty} I_n$ contains only one point.

Proof (*): Note that the sequences (a_n) and (b_n) are respectively increasing and decreasing sequences; moreover both are bounded. Hence both converge, say $a_n \to a$ and $b_n \to b$. Then $a_n \leq a$ and $b \leq b_n$ for all $n \in N$. Since $b - a = \lim(b_n - a_n) = 0$, $a = b$. Since $a_n \leq b_n$ for all n we have $a \in \bigcap\limits_{n=1}^{\infty} I_n$. Clearly if $x \neq a$ then x does not belong to $\bigcap\limits_{n=1}^{\infty} I_n$. $\square$

Subsequences : Let (x_n) be a sequence and let (n_k) be any sequence of positive integers such that $n_1 < n_2 < n_3 < \ldots$. *The sequence (x_{n_k}) is called a subsequence.* Note that here k varies from 1 to ∞.

A subsequence is formed by deleting some of the elements of the sequence and retaining the remaining in the same order. For example, $(\frac{1}{k^2})$ and $(\frac{1}{2^k})$ (k varies from 1 to ∞) are subsequences of $(\frac{1}{n})$, where $n_k = k^2$ and $n_k = 2^k$.

Sequences $(1, 1, 1, \ldots)$ and $(0, 0, 0, \ldots)$ are both subsequences of $(1, 0, 1, 0, \ldots)$. From this we see that a given sequence may have convergent subsequences though the sequence itself is not convergent. We note that every sequence is a subsequence of itself and if $x_n \to x$ then every subsequence of (x_n) also converges to x.

The following theorem which is an important result in calculus, is a consequence of the nested interval theorem.

Theorem 3.2 ***(Bolzano-Weierstrass theorem)*:** *Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof (*):*(Sketch).* Let (x_n) be a bounded sequence such that the set $\{x_1, x_2, \cdots\} \subset [a, b]$. Divide this interval into two equal parts. Let I_1 be that interval which contains an infinite number of elements (or say terms) of (x_n). Let x_{n_1} be one of the elements belonging to the interval I_1. Divide I_1 into two equal parts and let I_2 be that interval which contains an infinite number of elements. Choose a point x_{n_2} in I_2 such that $n_2 > n_1$. Keep dividing the intervals I_k, to generate I_k's and x_{n_k}'s. By nested interval theorem $\bigcap\limits_{k=1}^{\infty} I_k = \{x\}$, for some $x \in [a, b]$. It is easy to see that the subsequence (x_{n_k}) converges to x. □

Theorem 3.3: *If a sequence (x_n) satisfies the Cauchy criterion then (x_n) converges.*

Proof (*): Let (x_n) satisfy the Cauchy criterion. Since (x_n) is bounded, by the previous theorem there exists a subsequence (x_{n_k}) convergent to some x_0. We now show that $x_n \to x_0$. Let $\epsilon > 0$. Since (x_n) satisfies the Cauchy criterion,

$$\text{there exists } N_1 \ s.t. \ \mid x_n - x_m \mid \leq \ \epsilon/2 \quad \text{for all } n, m \ \geq \ N_1 \ \ ...... \ (1)$$

Since $x_{n_k} \longrightarrow x_0$,

$$\text{there exists } \ N_2 \ s.t. \ \mid x_{n_k} - x_0 \mid \leq \ \epsilon/2 \ \text{ for all } \ n_k \ \geq \ N_2 \quad .... \ (2)$$

Let $N = max\{N_1, N_2\}$. For $n \geq N$, choose some $n_k \geq N$, then by (1) and (2) we have

$$\mid x_n - x_0 \mid \leq \mid x_n - x_{n_k} \mid + \mid x_{n_k} - x_0 \mid \leq \ \epsilon/2 + \epsilon/2 = \epsilon.$$

This proves that $x_n \to 0$. □

Checking the Cauchy criterion directly from the definition is very difficult. The following result will help us to check the Cauchy criterion.

Problem 3.4: *Suppose $0 < \alpha < 1$ and (x_n) is a sequence satisfying the* ***contractive condition****:*

$$|x_{n+2} - x_{n+1}| \ \leq \ \alpha|x_{n+1} - x_n| \qquad n = 1, 2, 3, \ldots.$$

Then show that (x_n) satisfies the Cauchy criterion.

Solution : Note that $|x_{n+2} - x_{n+1}| \leq \alpha|x_{n+1} - x_n| \leq \alpha^2|x_n - x_{n-1}| \leq \cdots \leq \alpha^n|x_2 - x_1|$.

For $n > m$, $|x_n - x_m| \leq (\alpha^{n-2} + \alpha^{n-3} + \cdots + \alpha^{m-1})|x_2 - x_1| \leq \frac{\alpha^m}{1-\alpha}|x_2 - x_1| \to 0$ as $m \to \infty$.

Thus (x_n) satisfies the Cauchy criterion. □

Examples 3.5: 1. Let $x_1 = 1$ and $x_{n+1} = \frac{1}{2+x_n}$. Then

$$\mid x_{n+2} - x_{n+1} \mid = \frac{1}{(2 + x_{n+1})(2 + x_n)} \mid x_n - x_{n+1} \mid < \ \frac{1}{4} \mid x_n - x_{n+1} \mid .$$

Therefore (x_n) satisfies the contractive condition with $\alpha = 1/4$ and hence it satisfies the Cauchy criterion. Therefore it converges. Suppose $x_n \to l$. Then $l = \frac{1}{2+l}$. Find l !.

Remark : *Whenever we use the result given in the above exercise, we have to show that the number α that we get, satisfies $0 < \alpha < 1$.*

2. If $x_1 = 2$ and $x_{n+1} = 2 + \frac{1}{x_n}$ then $\mid x_{n+2} - x_{n+1} \mid < \ \frac{1}{4} \mid x_n - x_{n+1} \mid$ (verify !). Therefore the sequence (x_n) converges.

Lecture 4 : Continuity and limits

Intuitively, we think of a function $f : \mathbb{R} \to \mathbb{R}$ as continuous if it has a continuous curve. The term *continuous curve* means that the graph of f can be drawn without *jumps*, i.e., the graph can be drawn with a *continuous* motion of the pencil without leaving the paper.

Suppose a function $f : \mathbb{R} \to \mathbb{R}$ has a discontinuous graph as shown in the following figure.

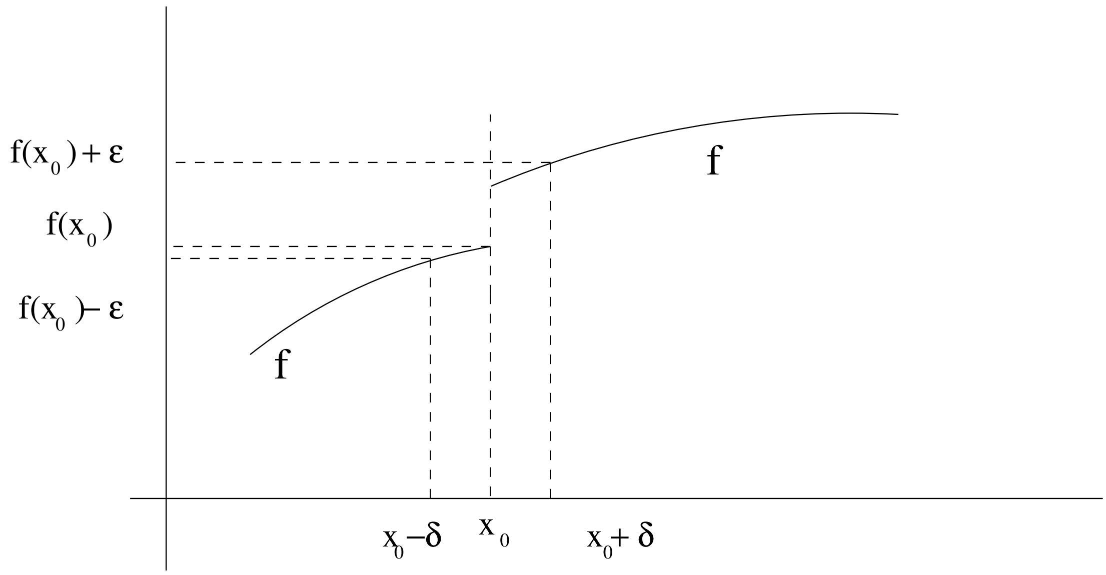


Figure 1: Discontinuous Graph

The graph is broken at the point $(x_0, f(x_0))$, i.e., the function f is discontinuous at x_0. Hence whenever x is close to x_0 from the right, $f(x)$ does not get close to $f(x_0)$. (The idea of getting close has already been discussed while dealing with convergent sequences). As shown in the figure, we can choose a neighbourhood $(f(x_0) - \epsilon_0, f(x_0) + \epsilon_0)$, $\epsilon_0 > 0$, at $f(x_0)$ such that if we take **any** neighbourhood $(x_0 - \delta, x_0 + \delta)$, $\delta > 0$, then the image of the interval $(x_0 - \delta, x_0 + \delta)$ does not lie inside $(f(x_0) - \epsilon_0, f(x_0) + \epsilon_0)$. In formal terms, **there exists** $\epsilon > 0$ such that **for all** $\delta > 0$, $|x - x_0| < \delta \not\Rightarrow |f(x) - f(x_0) < \epsilon$. Hence if a function f is not continuous at x_0, we have the above condition.

We will now give the formal definition of continuity of a function at a point (in the "ϵ-δ language").

Definition A function $f : \mathbb{R} \to \mathbb{R}$ is said to be continuous at a point $x_0 \in \mathbb{R}$ if for every $\epsilon > 0$, there is a $\delta > 0$ such that $|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_o| < \delta$.

Using the (visible) discontinuity in the above example, we were able to find some ϵ for which it was not possible to find any δ as in the definition. Roughly, f is continuous at x_0 if whenever x approaches x_0, $f(x)$ approaches $f(x_0)$. In some cases when f is not continuous at x_0, there may be a number A such that whenever x approaches x_0, $f(x)$ approaches A. In this case we call such a number A the limit of f at x_0. Formally, we have:

Definition : A number A is called the limit of a function f at a point x_0 if for every $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - A| < \epsilon$ whenever $0 < |x - x_0| < \delta$. If such a number A exists then it is unique.

In this case we write $\lim\limits_{x \to x_0} f(x) = A$. It is clear that $f(x_0)$ is the limit of f at x_0 if f is continuous at x_0.

The reader is advised to see the strong analogy between the definition of limit point and the definition of convergence of sequence. Let us now characterize the continuity of a function at a point in terms of sequences.

Theorem 4.1 : *A real valued function f is continuous at $x_0 \in \mathbb{R}$ if and only if whenever a sequence of real numbers (x_n) converges to x_0, then the sequence $(f(x_n))$ converges to $f(x_0)$.*

Proof: Suppose f is continuous at x_0 and $x_n \to x_0$. Let us show that $f(x_n) \to f(x_0)$. Let $\epsilon > 0$ be given. We must find N such that $|f(x_n) - f(x_0)| < \epsilon$ for all $n \geq N$. Since f is continuous at x_0, there exists $\delta > 0$ such that $|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_0| < \delta$. Since $x_n \to x_0$, there exists N such that $|x_n - x_0| < \delta$ for all $n \geq N$. This N serves our purpose.

To prove the converse, let us assume the contrary that f is not continuous at x_0. Then for some $\epsilon > 0$ and for each n, there is an element x_n such that $|x_n - x_0| < \frac{1}{n}$ but $|f(x_n) - f(x_0)| \geq \epsilon$. This contradicts the fact that $x_n \to x_0$ implies $f(x_n) \to f(x_0)$. $\square$

Remark : To define the continuity of a function f at a point x_0, the function f has to be defined at x_0. But even if the function is not defined at x_0, one can define the limit of a function at x_0.

The proof of the following theorem is similar to the proof of the previous theorem.

Theorem 4.2: $\lim\limits_{x \to x_0} f(x) = A$ *if and only if whenever a sequence of real numbers (x_n) converges to x_0, $x_n \neq x_0$ for all n, then the sequence $(f(x_n))$ converges to A.*

Examples : 1. Define a function $f(x)$ such that $f(x) = 2xsin(\frac{1}{x})$ when $x \neq 0$ and $f(0) = 0$. We will show that f is continuous at 0 using first by the $\varepsilon - \delta$ definition and then by the sequential characterization.

Using the $\varepsilon - \delta$ definition : Remember that for a given $\varepsilon > 0$, we have to find a $\delta > 0$ (not the other way!). Note that here $x_0 = 0$ and

$$\mid f(x) - f(x_0) \mid = \mid 2xsin(\frac{1}{x}) - 0 \mid \leq \mid 2x \mid = 2 \mid x - x_0 \mid .$$

Suppose that ε is given. Choose any $\delta > 0$ such that $\delta \leq \frac{\varepsilon}{2}$. Then we have

$$\mid f(x) - f(x_0) \mid < \varepsilon \text{ whenever } \quad \mid x - x_0 \mid < \delta.$$

This shows that f is continuous at $x_0 = 0$.

Using the sequential characterization : Note that $\mid f(x) \mid \leq 2 \mid x \mid$. Therefore, $f(x_n) \to f(0)$ whenever $x_n \to 0$. This proves that f is continuous at 0.

2. The function $f(x) = \sin(1/x)$ is defined for all $x \neq 0$. This function has no limit as $x \to 0$ because if we take $x_n = 2/\{\pi(2n+1)\}$ for $n = 1, 2, \ldots$, then $x_n \to 0$ but $f(x_n) = (-1)^n$ which does not tend to any limit as $n \to \infty$.

3. Let $f(x) = 0$ when x is rational and $f(x) = x$ when x is irrational. We will see that this function is continuous only at $x = 0$. Let (x_n) be any sequence such that $x_n \to 0$. Because, $\mid f(x_n) \mid \leq \mid x_n \mid$, $f(x_n) \to f(0)$. Therefore f is continuous at 0.

Suppose $x_0 \neq 0$ and it is rational. We will show that f is not continuous at x_0. Choose (x_n) such that $x_n \to x_0$ and all $x_n's$ are irrational numbers. Then $f(x_n) = x_n \to x_0 \neq f(x_0)$.This proves that f is not continuous at x_0. When x_0 is irrational, the proof is similar.

Remark : In order to show that a function is not continuous at a point x_0 it is sufficient to produce one sequence (x_n) such that $x_n \to x_0$ but $f(x_n) \nrightarrow f(x_0)$. However, to show a function is continuous at x_0, we have to show that $f(x_n) \to f(x_0)$ whenever $x_n \to x_0$ i.e, for every (x_n) such that $x_n \to x_0$.

Continuous function on a subset of $\mathbb{R}$: Let S be a subset of $\mathbb{R}$ and $x_0 \in S$, we say that f is continuous at x_0, if for every $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $x \in S$ with $|x - x_0| < \delta$ we have $|f(x) - f(x_0)| < \epsilon$. Moreover, if f is continuous at each $x \in S$, then we say that f is continuous on S.

Limits at Infinity : Let $f : \mathbb{R} \to \mathbb{R}$. We say that $\lim\limits_{x\to\infty} f(x) = A$ if for every $\epsilon > 0$, there exist $N > 0$ such that whenever $x \geq N$, we have $|f(x) - A| < \epsilon$.

Let $x_0 \in \mathbb{R}$. We say that $\lim\limits_{x\to x_0} f(x) = \infty$ if for every M, there exists $\delta > 0$ such that whenever $|x - x_0| < \delta$ we have $f(x) > M$.

Problem 1: *Let $f : \mathbb{R} \to \mathbb{R}$ be such that for every $x, y \in \mathbb{R}$, $\mid f(x) - f(y) \mid \leq \mid x - y \mid$. Show that f is continuous.*

Solution : Let $x_0 \in \mathbb{R}$ and $x_n \to x_0$. Since $\mid f(x_n) - f(x_0) \mid \leq \mid x_n - x_0 \mid, f(x_n) \to f(x_0)$. Therefore f is continuous at x_0. Since x_0 is arbitrary, f is continuous everywhere.

Problem 2: *Let $f : (-1, 1) \to \mathbb{R}$ be a continuous function such that in every neighborhood of 0, there exists a point where f takes the value 0. Show that $f(0) = 0$.*

Solution : For every n, there exists $x_n \in (-\frac{1}{n}, \frac{1}{n})$ such that $f(x_n) = 0$. Since f is continuous at 0 and $x_n \to 0$, we have $f(x_n) \to f(0)$. Therefore, $f(0) = 0$.

Problem 3: *Let $f : \mathbb{R} \to \mathbb{R}$ satisfy $f(x + y) = f(x) + f(y)$ for all $x, y \in \mathbb{R}$. If f is continuous at 0, show that f is continuous at every point $c \in \mathbb{R}$.*

Solution : First note that $f(0) = 0, f(-x) = -f(x)$ and $f(x - y) = f(x) - f(y)$. Let $x_0 \in \mathbb{R}$ and $x_n \to x_0$. Then $f(x_n) - f(x_0) = f(x_n - x_0) \to f(0) = 0$ as f is continuous at 0 and $x_n - x_0 \to 0$.

Properties of Continuous Functions on a Closed Interval :

Definition : Let $S \subseteq \mathbb{R}$ and $f : S \to \mathbb{R}$. We say that f is bounded on S if the set $f(S) := \{f(x) : x \in S\}$ is a bounded subset of $\mathbb{R}$.

We will now see some properties of continuous functions on a closed interval.

Theorem 4.3 : *If a function f is continuous on $[a, b]$ then it is bounded on $[a, b]$.*

Proof: Suppose that f is not bounded on $[a, b]$. Then for each natural number n there is a point $x_n \in [a, b]$ such that $|f(x_n)| > n$. Since (x_n) is a bounded sequence, by Bolzano-Weierstrass theorem it has a convergent subsequence, say $x_{n_k} \to x_0 \in [a, b]$. By the continuity of f, we have $f(x_{n_k}) \to f(x_0)$. This contradicts the assumption that $|f(x_n)| > n$ for all n. Hence f is bounded on $[a, b]$. $\square$

We remark that if a function is continuous on an open interval (a, b) or on a semi-open interval of the type $(a, b]$ or $[a, b)$, then it is not necessary that the function has to be bounded. For example, consider the continuous function $\frac{1}{x}$ on $(0, 1]$.

Lecture 5 : Existence of Maxima, Intermediate Value Property, Differentiabilty

Let f be defined on a subset S of $\mathbb{R}$. An element $x_0 \in S$ is called a *maximum* for f on S if $f(x_0) \geq f(x)$ for all $x \in S$ and in this case $f(x_0)$ is the maximum value f. Similarly, x_0 is called a *minimum* for f on S if $f(x_0) \leq f(x)$ for all $x \in S$.

The following theorem provides an existence criterion for maximum and minimum.

Theorem 5.1: *Let f be continuous on $[a, b]$. Then there exist $x_0, y_0 \in [a, b]$ such that x_0 is a maximum for f on $[a, b]$ and y_0 is a minimum for f on $[a, b]$.*

Proof (*): By Theorem 4.3, f is bounded on $[a, b]$. Let M be the least upper bound of $f([a, b])$ ($:= \{f(x) : x \in [a, b]\}$). Then there exists a sequence $\{f(x_n)\}$ in $f([a, b])$ such that $f(x_n) \to M$. Since $\{x_n\}$ is a bounded sequence in $[a, b]$, it has a convergent subsequence, say $x_{n_k} \to x_0 \in [a, b]$. By the continuity of f we have $f(x_0) = M$. Hence x_0 is a maximum for f on $[a, b]$.

The proof for the existence of a minimum is similar. □

Consider a function $f : [-1, 1] \to \mathbb{R}$ such that f is continuous and satisfies $f(-1) < 0$ and $f(1) > 0$. Intuitively, we feel that the graph of f should cross the x-axis between -1 and 1; i.e., there is some $x_0 \in [-1, 1]$ such that $f(x_0) = 0$. This motivates us to state the following theorem.

Theorem 5.2 (Intermediate Value Property) : *Let f be continuous on $[a, b]$, and let $f(a) < s < f(b)$ (s is a value which is intermediate between two values taken by f). Then there exists x such that $a < x < b$ and $f(x) = s$.*

Proof (*): Let $S = \{x \in [a, b] : f(x) \leq s\}$. Since $a \in S$, we have $S \neq \emptyset$ and S is bounded above by b. Let c be the least upper bound of S. We claim that $f(c) = s$. Since c is the least upper bound of S, there exists a sequence $\{x_n\}$ from S such that $x_n \to c$. By the continuity of f, $f(x_n) \to f(c)$. Since $f(x_n) \leq s$ for all n, we have $f(c) \leq s$. Note that $b > c$. Consider the sequence $y_n = c + (b - c)/n$. As $y_n \to c$, we get $f(y_n) \to f(c)$. Since $f(y_n) > s$ for all n, we have $f(c) \geq s$. It follows that $f(c) = s$. □

Problem 1 : *Show that the equation $(1 - x)cosx = sinx$ has at least one solution in $(0, 1)$.*

Solution: Set $f(x) = (1-x)cosx - sinx$. Then $f(0) = 1$ and $f(1) = -sin1 < 0$. By the intermediate value property there is $x_0 \in (0, 1)$ such that $f(x_0) = 0$.

Problem 2 : *Let $f : [0, 1] \to [0, 1]$ be continuous. Show that f has a fixed point in $[0, 1]$; that is, there exists $x_0 \in [0, 1]$ such that $f(x_0) = x_0$.*

Solution: Define the function $g(x) = f(x) - x$ on $[0, 1]$. Then f is continuous, $g(0) \geq 0$ and $g(1) \leq 0$. Use the intermediate value property (IVP).

Problem 3 : *Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Show that the range $\{f(x) : x \in [a, b]\}$ is a closed and bounded interval.*

Solution: Since f is a continuous function, there exist $x_0, y_0 \in [a, b]$ such that $f(x_0) = m = inf f$ and $f(y_0) = M = sup f$. Suppose $x_0 < y_0$. By the IVP, for every $\alpha \in [m, M]$ there exists $x \in [x_0, y_0]$ such that $f(x) = \alpha$. Hence $f([a, b]) = [m, M]$.

Problem 4 : *Show that a polynomial of odd degree has at least one real* root.

Solution: Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, $a_n \neq 0$ and n be odd. Then $p(x) =$

$x^n(a_n + \frac{a_{n-1}}{x} + \cdots + \frac{a_1}{x^{n-1}} + \frac{a_0}{x^n})$. If $a_n > 0$, then $p(x) \to \infty$ as $x \to \infty$ and $p(x) \to -\infty$ as $x \to -\infty$. Thus by the IVP, there exists x_0 such that $p(x_0) = 0$. Similar argument for $a_n < 0$.

Differentiation : We now deal with derivatives, an important concept of differential calculus. The reader must be familiar with this from elementary calculus. For example, the geometric problem of finding the tangent line to a curve at a given point leads to the notion of derivative.

Definition : Let I be an interval which is not a singleton and let f be a function defined on I. A function f is said to be differentiable at $x \in I$ if the limit

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

exists.

It is understood that the limit is taken for $x + h \in I$; thus if x is a left end point of I then we only consider $h > 0$. If the above limit exists, it is called the derivative of f at x and is denoted by $f'(x)$. If f is differentiable at each $x \in I$, then f' is a function on I.

It is clear that if f is differentiable at $c \in I$, then

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

Now we prove that differentiability implies continuity.

Theorem 5.3 : *Let f be defined on an interval I. If f is differentiable at a point $c \in I$, then f is continuous at c.*

Proof: As $x \to c$, $x \neq c$, we have $f(x) - f(c) = \frac{f(x)-f(c)}{x-c}\ (x - c) \ \to\ f'(x) \cdot 0 = 0.$ ☐

Problem 5: *Show that the function $f(x)$ defined by $f(x) = x^2 sin\frac{1}{x}$, when $x \neq 0$ and $f(0) = 0$ is differentiable at all $x \in \mathbb{R}$. Also show that the function $f'(x)$ is not continuous at $x = 0$.*

Solution : Note that $f'(0) = \lim_{x\to 0} \frac{x^2 sin\frac{1}{x}}{x} = 0$. At other points the function is differentiable because the function is a product of two differentiable functions (here note that $sin\frac{1}{x}$ is a composition of two functions) and $f'(x) = 2x\sin\frac{1}{x} - \cos\frac{1}{x}$, for $x \neq 0$. Since $\lim\limits_{h\to 0} \cos\frac{1}{h}$ does not exist, $f'(x)$ is not continuous at 0.

Problem 6: *If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $c \in \mathbb{R}$, show that $f'(c) = lim(n\{f(c+1/n) - f(c)\})$. However, show by example that the existence of the limit of this sequence does not imply the existence of $f'(c)$.*

Solution : Since $f'(c)$ exists, by taking $h = 1/n$ in the definition of the differentiability we get that $f'(c) = lim(n\{f(c+1/n) - f(c)\})$.

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$ for x rational, $f(x) = 1$ for x irrational. Then the function is not even continuous at 0; hence it cannot be differentiable at 0. However, $lim(n\{f(c+1/n) - f(c)\})$ exists at $c = 0$.

Problem 7: *Let $f(0) = 0$ and $f'(0) = 1$. For a positive integer k, show that*

$$\lim_{x\to 0} \frac{1}{x}\left\{f(x) + f(\frac{x}{2}) + f(\frac{x}{3}) + ... + f(\frac{x}{k})\right\} = 1 + \frac{1}{2} + ... + \frac{1}{k}.$$

Solution : $\lim\limits_{x\to 0} \frac{1}{x}(f(x) + f(\frac{x}{2}) + f(\frac{x}{3}) + \cdots + f(\frac{x}{k})) =$

$$\lim_{x\to 0}(\frac{f(x)-f(0)}{x} + \frac{1}{2}\frac{f(\frac{x}{2})-f(0)}{\frac{x}{2}} + \cdots + \frac{1}{k}\frac{f(\frac{x}{k})-f(0)}{\frac{x}{k}}) = 1 + \frac{1}{2} + \cdots + \frac{1}{k}.$$

Lecture 6 : Rolle's Theorem, Mean Value Theorem

The reader must be familiar with the classical maxima and minima problems from calculus. For example, the graph of a differentiable function has a horizontal tangent at a maximum or minimum point. This is not quite accurate as we will see.

Definition : Let $f : I \to \mathbb{R}$, I an interval. A point $x_0 \in I$ is a local maximum of f if there is a $\delta > 0$ such that $f(x) \leq f(x_0)$ whenever $x \in I \cap (x_0 - \delta, x_0 + \delta)$. Similarly, we can define local minimum.

Theorem 6.1 : *Suppose $f : [a, b] \to \mathbb{R}$ and suppose f has either a local maximum or a local minimum at $x_0 \in (a, b)$. If f is differentiable at x_0 then $f'(x_0) = 0$.*

Proof: Suppose f has a local maximum at $x_0 \in (a, b)$. For small (enough) h, $f(x_0 + h) \leq f(x_0)$. If $h > 0$ then

$$\frac{f(x_0 + h) - f(x_0)}{h} \leq 0.$$

Similarly, if $h < 0$, then

$$\frac{f(x_0 + h) - f(x_0)}{h} \geq 0.$$

By elementary properties of the limit, it follows that $f'(x_0) = 0$. $\square$

We remark that the previous theorem is not valid if x_0 is a or b. For example, if we consider the function $f : [0, 1] \to \mathbb{R}$ such that $f(x) = x$, then f has maximum at 1 but $f'(x) = 1$ for all $x \in [0, 1]$.

The following theorem is known as *Rolle's theorem* which is an application of the previous theorem.

Theorem 6.2 : *Let f be continuous on $[a, b]$, $a < b$, and differentiable on (a, b). Suppose $f(a) = f(b)$. Then there exists c such that $c \in (a, b)$ and $f'(c) = 0$.*

Proof: If f is constant on $[a, b]$ then $f'(c) = 0$ for all $c \in [a, b]$. Suppose there exists $x \in (a, b)$ such that $f(x) > f(a)$. (A similar argument can be given if $f(x) < f(a)$). Then there exists $c \in (a, b)$ such that $f(c)$ is a maximum. Hence by the previous theorem, we have $f'(c) = 0$. $\square$

Problem 1 : *Show that the equation $x^{13} + 7x^3 - 5 = 0$ has exactly one (real) root.*

Solution : Let $f(x) = x^{13} + 7x^3 - 5$. Then $f(0) < 0$ and $f(1) > 0$. By the IVP there is at least one positive root of $f(x) = 0$. If there are two distinct positive roots, then by Rolle's theorem there is some $x_0 > 0$ such that $f'(x_0) = 0$ which is not true. Moreover, observe that $f(x) < 0$ for $x < 0$.

Problem 2 : *Let f and g be functions, continuous on $[a, b]$, differentiable on (a, b) and let $f(a) = f(b) = 0$. Prove that there is a point $c \in (a,\ b)$ such that $g'(c)f(c) + f'(c) = 0$.*

Solution : Define $h(x) = f(x)e^{g(x)}$. Here, $h(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Since $h(a) = h(b) = 0$, by Rolle's theorem, there exists $c \in (a, b)$ such that $h'(c) = 0$.

Since $h'(x) = [f'(x) + g'(x)f(x)]e^{g(x)}$ and $e^{\alpha} \neq 0$ for any $\alpha \in \mathbb{R}$, we see that $f'(c) + g'(c)f(c) = 0$.

A geometric interpretation of the above theorem can be given as follows. If the values of a differentiable function f at the end points a and b are equal then somewhere between a and b there is a horizontal tangent. It is natural to ask the following question. If the value of f at the end points a and b are not the same, is it true that there is some $c \in [a, b]$ such that the tangent line at c is parallel to the line connecting the endpoints of the curve? The answer is yes and this is essentially the Mean Value Theorem.

Theorem 6.3 : (Mean Value Theorem) *Let f be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $c \in (a,b)$ such that $f(b) - f(a) = f'(c)(b-a)$.*

Proof: Let

$$g(x) = f(x) - \frac{f(b) - f(a)}{b-a}\,(x-a).$$

Then $g(a) = g(b) = f(a)$. The result follows by applying Rolle's Theorem to g. □

The mean value theorem is an important result in calculus and has some important applications relating the behaviour of f and f'. For example, if we have a property of f' and we want to see the effect of this property on f, we usually try to apply the mean value theorem. Let us see some examples.

Example 1 : *Let $f : [a,b] \to \mathbb{R}$ be differentiable. Then f is constant if and only if $f'(x) = 0$ for every $x \in [a,b]$.*

Proof : Suppose that f is constant, then from the definition of $f'(x)$ it is immediate that $f'(x) = 0$ for every $x \in [a,b]$.

To prove the converse, let $a < x \leq b$. By the mean value theorem there exists $c \in (a,x)$ such that $f(x) - f(a) = f'(c)(x-a)$. Since $f'(c) = 0$, we conclude that $f(x) = f(a)$, that is f is constant. (*If we try to prove the converse directly from the definition of $f'(x)$ we will be in trouble.*) □

Example 2 : *Suppose f is continuous on $[a,b]$ and differentiable on (a,b).*

(i) If $f'(x) \neq 0$ for all $x \in (a,b)$, then f is one-one (i.e, $f(x) \neq f(y)$ whenever $x \neq y$).

(ii) If $f'(x) \geq 0$ (resp. $f'(x) > 0$) for all $x \in (a,b)$ then f is increasing (resp. strictly increasing) on $[a,b]$. (We have a similar result for decreasing functions.)

Proof : Apply the mean value theorem as we did in the previous example. (Note that f can be one-one but f' can be 0 at some point, for example take $f(x) = x^3$ and $x = 0$.)

Problem 3 : *Use the mean value theorem to prove that $\mid sinx - siny \mid \leq \mid x - y \mid$ for all $x, y \in \mathbb{R}$.*

Solution : Let $x, y \in \mathbb{R}$. By the mean value theorem $sinx - siny = cosc\ (x-y)$ for some c between x and y. Hence $\mid sinx - siny \mid \leq \mid x - y \mid$.

Problem 4 : *Let f be twice differentiable on $[0,2]$. Show that if $f(0) = 0, f(1) = 2$ and $f(2) = 4$, then there is $x_0 \in (0,2)$ such that $f''(x_0) = 0$.*

Solution : By the mean value theorem there exist $x_1 \in (0,1)$ and $x_2 \in (1,2)$ such that

$$f'(x_1) = f(1) - f(0) = 2 \quad \text{and} \quad f'(x_2) = f(2) - f(1) = 2.$$

Apply Rolle's theorem to f' on $[x_1, x_2]$.

Problem 5 : *Let $a > 0$ and $f : [-a,a] \to \mathbb{R}$ be continuous. Suppose $f'(x)$ exists and $f'(x) \leq 1$ for all $x \in (-a,a)$. If $f(a) = a$ and $f(-a) = -a$, then show that $f(x) = x$ for every $x \in (-a,a)$.*

Solution : Let $g(x) = f(x) - x$ on $[-a,a]$. Note that $g'(x) \leq 0$ on $(-a,a)$. Therefore, g is decreasing. Since $g(a) = g(-a) = 0$, we have $g = 0$.

This problem can also be solved by applying the MVT for g on $[-a,x]$ and $[x,a]$.

Lecture 7 : Cauchy Mean Value Theorem, L'Hospital Rule

L'Hospital (pronounced Lopeetal) Rule is a useful method for finding limits of functions. There are several versions or forms of L'Hospital rule. Let us start with one form called $\frac{0}{0}$ form which deals with $\lim_{x\to x_0} \frac{f(x)}{g(x)}$, where $\lim_{x\to x_0} f(x) = 0 = \lim_{x\to x_0} g(x)$.

Theorem 1: (L'Hospital Rule) *Let $f, g : (a,b) \to \mathbb{R}$ be differentiable at $x_0 \in (a,b)$. Suppose $f(x_0) = g(x_0) = 0$ and $g'(x_0) \neq 0$. Then $\lim_{x\to x_0} \frac{f(x)}{g(x)} = \frac{f'(x_0)}{g'(x_0)}$.*

Proof : Note that $\frac{f'(x_0)}{g'(x_0)} = \frac{\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}}{\lim_{x\to x_0} \frac{g(x)-g(x_0)}{x-x_0}} = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{g(x)-g(x_0)} = \lim_{x\to x_0} \frac{f(x)}{g(x)}$. □

Example : The condition $f(x_0) = g(x_0) = 0$ is essential in the previous result. For example, $\lim_{x\to 0} \frac{x+17}{2x+3} = \frac{17}{3}$ but $\frac{f'(0)}{g'(0)} = \frac{1}{2}$.

The following result is a stronger version of the previous result.

Theorem 2 : (L'Hospital Rule) *Let $f, g : [x_0, b) \to \mathbb{R}$. Suppose $f(x_0) = g(x_0) = 0$ and f, g are differentiable on (x_0, b). Let $g'(x) \neq 0$ for all $x \in (x_0, b)$. Then*

$$\lim_{x\to x_0^+} \frac{f(x)}{g(x)} = \lim_{x\to x_0^+} \frac{f'(x)}{g'(x)}$$

provided $\lim_{x\to x_0^+} \frac{f'(x)}{g'(x)}$ exists.

We note that the rule given in Theorem 2 is also valid for left hand limits with similar assumptions on f and g.

The difference between Theorem 1 and Theorem 2 is that in Theorem 2 the function g may not be differentiable at x_0 but this is not the case in Theorem 1. So in order to prove Theorem 2, we have to modify the technique used in the proof of Theorem 1. Basically we have to handle the quotient $\frac{f(x)-f(x_0)}{g(x)-g(x_0)}$ appearing in the proof of Theorem 1 in a different way. For this, we need the following theorem.

Theorem 3 : (Cauchy Mean Value Theorem) *Let f and g be continuous on $[a,b]$ and differentiable on (a,b). Suppose that $g'(x) \neq 0$ for all $x \in (a,b)$. Then there exists $c \in (a,b)$ such that*

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

Proof (*): Consider the function

$$F(x) = f(x) - f(a) - \frac{f(b)-f(a)}{g(b)-g(a)}(g(x) - g(a)).$$

Then F is continuous on $[a,b]$, differentiable on (a,b) and $F(a) = F(b) = 0$. By Rolle's Theorem there exists $c \in (a,b)$ such that $F'(c) = 0$. This proves the theorem. □

Remark : Cauchy mean value theorem (CMVT) is sometimes called generalized mean value theorem. Because, if we take $g(x) = x$ in CMVT we obtain the MVT. We will use CMVT to prove Theorem 2. We will now see an application of CMVT.

Problem 1: *Using Cauchy Mean Value Theorem, show that $1 - \frac{x^2}{2!} < \cos x$ for $x \neq 0$.*

Solution: Apply CMVT to $f(x) = 1 - \cos x$ and $g(x) = \frac{x^2}{2}$. We get $\frac{1-\cos x}{x^2/2} = \frac{\sin c}{c} < 1$ for some c between 0 and x.

Problem 2: *Let f be continuous on $[a, b]$, $a > 0$ and differentiable on (a, b). Prove that there exists $c \in (a, b)$ such that $\frac{bf(a)-af(b)}{b-a} = f(c) - cf'(c)$.*

Solution: Apply CMVT to $\frac{f(x)}{x}$ and $\frac{1}{x}$.

Proof of Theorem 2(*): We will show that if $\lim_{x\to x_0^+} \frac{f'(x)}{g'(x)} = l$ then $\lim_{x\to x_0^+} \frac{f(x)}{g(x)} = l$. Suppose $x_n \to x_0{}^+$. By CMVT (on $[x_0, x_n]$), there exists $c_n \in (x_0, x_n)$ such that

$$\frac{f'(c_n)}{g'(c_n)} = \frac{f(x_n) - f(x_0)}{g(x_n) - g(x_0)} = \frac{f(x_n)}{g(x_n)}.$$

Therefore, $\lim_{x_n\to x_0^+} \frac{f(x_n)}{g(x_n)} = \lim_{c_n\to x_0^+} \frac{f'(c_n)}{g'(c_n)} = l$. □

Remarks : 1. In Theorem 2, we can also take $l = \infty$ or $l = -\infty$ and in this case the proof is essentially the same.

2. We can replace the condition $f(x_0) = g(x_0) = 0$ by the condition $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = 0$ in Theorem 2. In this case the functions f and g may not be defined at x_0. For the proof, extend the functions to $[x_0, b)$ such that $f(x_0) = g(x_0) = 0$.

3. Theorem 2 is also true for $x_0 = \infty$ or $-\infty$ i.e., if $\lim_{x\to\infty} \frac{f'(x)}{g'(x)} = l$ then $\lim_{x\to\infty} \frac{f(x)}{g(x)} = l$ with appropriate assumptions on f and g. To prove this fact, consider the functions $F(x) = f(\frac{1}{x})$ and $G(x) = g(\frac{1}{x}), x \neq 0$ and allow $x \to 0^-$. In this case also we can take $l = \infty$ or $l = -\infty$.

4. In case of the $\frac{\infty}{\infty}$ form we have

$$\{\lim_{x\to x_0^+} f(x) = \pm\infty = \lim_{x\to x_0^+} g(x)\} \quad \Rightarrow \quad \{\text{if } \lim_{x\to x_0^+} \frac{f'(x)}{g'(x)} = l \text{ then } \lim_{x\to x_0^+} \frac{f(x)}{g(x)} = l\}.$$

Of course we assume the standard necessary conditions as above on f and g. This result is also true when $x_0 = \infty$ or $-\infty$ and/or $l = \infty$ or $l = -\infty$. We will not give the proofs of these results but we will use them.

5. Other forms such as $\infty - \infty, 0(\infty), 1^\infty, ..$ are usually reduced to the form $\frac{0}{0}$ or $\frac{\infty}{\infty}$.

Examples : 1. $\lim_{x\to 0^+} \frac{log(sinx)}{logx} = \lim_{x\to 0^+} \frac{cosx/sinx}{1/x} = 1$ ($\frac{\infty}{\infty}$ form).

2. $\lim_{x\to 0^+}(xlogx) = \lim_{x\to 0^+} \frac{logx}{1/x} = \lim_{x\to 0^+} \frac{1/x}{-1/x^2} = 0$. ($0 \cdot \infty$ form).

3. $\lim_{x\to 0^+}(\frac{1}{x} - \frac{1}{sinx}) = \lim_{x\to 0^+} \frac{sinx-x}{xsinx} = \lim_{x\to 0^+} \frac{cosx-1}{sinx+xcosx} = \lim_{x\to 0^+} \frac{-sinx}{2cosx-xsinx} = 0$.

4. Let us evaluate $\lim_{x\to\infty}(1 + \frac{1}{x})^x = \lim_{x\to\infty} e^{xlog(1+\frac{1}{x})}$. Note that $xlog(1 + \frac{1}{x}) = \lim_{x\to\infty} \frac{log(1+\frac{1}{x})}{\frac{1}{x}} = \lim_{x\to\infty} \frac{1}{1+\frac{1}{x}} = 1$. Therefore, $\lim_{x\to\infty}(1 + \frac{1}{x})^x = e$.

5. Note that $\lim_{x\to\infty} \frac{x-sinx}{2x+sinx} = \frac{1}{2}$. However, one cannot apply L'Hospital Rule, because $\lim_{x\to\infty} \frac{1-cosx}{2+cosx}$ does not exist.

6. Note that $\lim_{x\to 0} \frac{x+1}{x}$ does not exist but if we apply L'Hospital rule, we get a wrong answer : $\lim_{x\to 0} \frac{x+1}{x} = \lim_{x\to 0} \frac{1}{1} = 1$. The reason is that in this case we cannot apply L'Hospital Rule because the given quotient is neither $\frac{0}{0}$ form nor $\frac{\infty}{\infty}$ form

Lecture 8 : Fixed Point Iteration Method, Newton's Method

In the previous two lectures we have seen some applications of the mean value theorem. We now see another application.

In this lecture we discuss the problem of finding approximate solutions of the equation

$$f(x) = 0. \tag{1}$$

In some cases it is possible to find the exact roots of the equation (1), for example, when $f(x)$ is a quadratic or cubic polynomial. Otherwise, in general, one is interested in finding approximate solutions using some (numerical) methods. Here, we will discuss a method called fixed point iteration method and a particular case of this method called Newton's method.

Fixed Point Iteration Method : In this method, we first rewrite the equation (1) in the form

$$x = g(x) \tag{2}$$

in such a way that any solution of the equation (2), which is a fixed point of g, is a solution of equation (1). Then consider the following algorithm.

Algorithm 1: Start from any point x_0 and consider the recursive process

$$x_{n+1} = g(x_n), \quad n = 0, 1, 2, ... \tag{3}$$

If f is continuous and (x_n) converges to some l_0 then it is clear that l_0 is a fixed point of g and hence it is a solution of the equation (1). Moreover, x_n (for a large n) can be considered as an approximate solution of the equation (1).

First let us illustrate whatever we said above with an example.

Example 1: We know that there is a solution for the equation $x^3 - 7x + 2 = 0$ in $[0, 1]$. We rewrite the equation in the form $x = \frac{1}{7}(x^3 + 2)$ and define the process $x_{n+1} = \frac{1}{7}(x_n^3 + 2)$. We have already seen in a tutorial class that if $0 \le x_0 \le 1$ then (x_n) satisfies the Cauchy criterion and hence it converges to a root of the above equation. We also note that if we start with (for example) $x_0 = 10$ then the recursive process does not converge.

It is clear from the above example that the convergence of the process (3) depends on g and the starting point x_0. Moreover, in general, showing the convergence of the sequence (x_n) obtained from the iterative process is not easy. So we ask the following question.

Question : Under what assumptions on g and x_0, does Algorithm 1 converge ? When does the sequence (x_n) obtained from the iterative process (3) converge ?

The following result is a consequence of the mean value theorem.

Theorem 8.1: *Let $g : [a, b] \to [a, b]$ be a differentiable function such that*

$$| g'(x) | \le \alpha < 1 \text{ for all } x \in [a, b]. \tag{4}$$

Then g has exactly one fixed point l_0 in $[a, b]$ and the sequence (x_n) defined by the process (3), with a starting point $x_0 \in [a, b]$, converges to l_0.

Proof (*): By the intermediate value property g has a fixed point, say l_0. The convergence of (x_n) to l_0 follows from the following inequalities:

$$| x_n - l_0 | = | g(x_{n-1}) - g(l_0) | \le \alpha | x_{n-1} - l_0 | \le \alpha^2 | x_{n-2} - l_0 | .... \le \alpha^n | x_0 - l_0 | \to 0.$$

If l_1 is a fixed point then $| l_0 - l_1 | = | g(l_0) - g(l_1) | \leq \alpha | l_0 - l_1 | < | l_0 - l_1 |$ This implies that $l_0 = l_1$. □

Example 2 : (i) Let us take the problem given in Example 1 where $g(x) = \frac{1}{7}(x^3 + 2)$. Then $g : [0,1] \to [0,1]$ and $| g'(x) | < \frac{3}{7}$ for all $x \in [0,1]$. Hence by the previous theorem the sequence (x_n) defined by the process $x_{n+1} = \frac{1}{7}(x_n^3 + 2)$ converges to a root of $x^3 - 7x + 2 = 0$.

(ii) Consider $f : [0,2] \to \mathbb{R}$ defined by $f(x) = (1 + x)^{1/5}$. Observe that f maps [0, 2] onto itself. Moreover $| f'(x) | \leq \frac{1}{5} < 1$ for $x \in [0,2]$. By the previous theorem the sequence (x_n) defined by $x_{n+1} = (1 + x_n)^{1/5}$ converges to a root of $x^5 - x - 1 = 0$ in the interval $[0,2]$.

In practice, it is often difficult to check the condition $f([a,b]) \subseteq [a,b]$ given in the previous theorem. We now present a variant of Theorem 1.

Theorem 8.2: *Let l_0 be a fixed point of $g(x)$. Suppose $g(x)$ is differentiable on $[l_0 - \varepsilon, l_0 + \varepsilon]$ for some $\varepsilon > 0$ and g satisfies the condition $| g'(x) | \leq \alpha < 1$ for all $x \in [l_0 - \varepsilon, l_0 + \varepsilon]$. Then the sequence (x_n) defined by (3), with a starting point $x_0 \in [l_0 - \varepsilon, l_0 + \varepsilon]$, converges to l_0.*

Proof : By the mean value theorem $g([l_0 - \varepsilon, l_0 + \varepsilon]) \subseteq [l_0 - \varepsilon, l_0 + \varepsilon]$ (Prove !). Therefore, the proof follows from the previous theorem. □

The previous theorem essentially says that if the starting point is sufficiently close to the fixed point then the chance of convergence of the iterative process is high.

Remark : If g is invertible then l_0 is a fixed point of g if and only if l_0 is a fixed point of g^{-1}. In view of this fact, sometimes we can apply the fixed point iteration method for g^{-1} instead of g. For understanding, consider $g(x) = 4x - 12$ then $| g'(x) | = 4$ for all x. So the fixed point iteration method may not work. However, $g^{-1}(x) = \frac{1}{4}x + 3$ and in this case $| (g^{-1})'(x) | = \frac{1}{4}$ for all x.

Newton's Method or Newton-Raphson Method :

The following iterative method used for solving the equation $f(x) = 0$ is called Newton's method.

Algorithm 2 : $\quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad n = 0, 1, 2, ...$

It is understood that here we assume all the necessary conditions so that x_n is well defined. If we take $g(x) = x - \frac{f(x)}{f'(x)}$ then Algorithm 2 is a particular case of Algorithm 1. So we will not get in to the convergence analysis of Algorithm 2. Instead, we will illustrate Algorithm 2 with an example.

Example 3: Suppose $f(x) = x^2 - 2$ and we look for the positive root of $f(x) = 0$. Since $f'(x) = 2x$, the iterative process of Newton's method is $x_{n+1} = \frac{1}{2}(x_n + \frac{2}{x_n}), n = 0, 1, 2, ...$ We have already discussed this sequence in a tutorial class. (Apparently, this process for calculating square roots was used in Mesopotamia before 1500 BC.)

Geometric interpretation of the iterative process of Newton's method : Suppose we have found $(x_n, f(x_n))$. To find x_{n+1}, we approximate the graph $y = f(x)$ near the point $(x_n, f(x_n))$ by the tangent : $y - f(x_n) = f'(x_n)(x - x_n)$. Note that x_{n+1} is the point of intersection of the x-axis and the tangent at x_n.

Lecture 9 : Sufficient Conditions for Local Maximum, Point of Inflection

In Lecture 6, we have seen a necessary condition for local maximum and local minimum. In this lecture we will see some sufficient conditions.

Sufficient Conditions for Local Maximum and Local Minimum

We will present sufficient conditions only for local maximum and the sufficient conditions for local minimum are similar. In the following results we assume $f : (a,b) \to \mathbb{R}$.

Theorem 9.1 : *Let $c \in (a,b)$ and f be continuous at c. If for some $\delta > 0$, f is increasing on $(c-\delta, c)$ and decreasing on $(c, c+\delta)$, then f has a local maximum at c.*

Proof : Choose any x_1 and x such that $c - \delta < x_1 < x < c$. Then $f(x_1) \leq f(x)$ and by the continuity of f at c we have

$$f(x_1) \leq \lim_{x \to c^-} f(x) = f(c).$$

Similarly, if $c < x_2 < c+\delta$ then $f(x_2) \leq \lim_{x\to c^+} f(x) = f(c)$. This proves the result. □

Corollary 9.1 : *Let $c \in (a,b)$ and f be continuous at c. If*

$$f'(x) \geq 0 \text{ for all } x \in (c-\delta, c) \text{ and } f'(x) \leq 0 \text{ for all } x \in (c, c+\delta)$$

then f has a local maximum at c.

Proof : The proof is immediate from the previous result. □

Corollary 9.2 : *Let $c \in (a,b)$. If $f'(c) = 0$ and $f''(c) < 0$ then f has a local maximum at c.*

Proof (*): Since $f''(c)$ exists, $f'(x)$ exists in a neighborhood of c. As $f''(c) < 0$, we have

$$f''(c) \;=\; \lim_{x\to c} \frac{f'(x) - f'(c)}{x-c} \;=\; \lim_{x \to c} \frac{f'(x)}{x-c} \;<\; 0.$$

Therefore there exists a $\delta > 0$ such that

$$\frac{f'(x)}{x-c} \;<\; 0 \text{ for all } x \in (c-\delta, c) \cup (c, c+\delta).$$

This implies that $f'(x) > 0$ for all $x \in (c-\delta, c)$ and $f'(x) < 0$ for all $x \in (c, c+\delta)$. Now apply the previous corollary. □

Remark : The converses of the previous results are not true, i.e.,

(i) If f is continuous at c and f has a local maximum at c, then f need not be increasing on $(c-\delta, c)$ or decreasing on $(c, c+\delta)$ for any $\delta > 0$. (Take the example : $f(x) = -(x sin(1/x))^2$ if $x \neq 0$, $f(0) = 0$ and $c = 0$.)

(ii) If f has a maximum at c and f is twice differentiable at c, then $f''(c)$ need not be less than 0. (Consider the example $f(x) = -x^4$ and $c = 0$).

So, the conditions assumed in the previous results are sufficient but not necessary.

Example : Let $f(x) = \frac{1}{x^4 - 2x^2 + 7} = \frac{1}{(x^2-1)^2 + 6}$, $x \in \mathbb{R}$. Then $f'(x) = \frac{-(4x^3 - 4x)}{(x^4 - 2x^2 + 7)^2} = \frac{-4x(x-1)(x+1)}{(x^4-2x^2+7)^2}$ and $f'(x) = 0$ for $x = -1, 0, 1$. By Corollary 9.1 and the corresponding result for local minimum,

f has a local minimum at $x = 0$ and local maxima at $x = -1$ and $x = 1$. In this example, it would be complicated to compute the second derivative and apply the second derivative test (Corollary 9.2).

Convexity, Concavity and Point of Inflection

Definition : Let $f : (a, b) \to \mathbb{R}$ be differentiable. We say that f is convex (resp., concave) on (a, b) if f' is strictly increasing (resp., strictly decreasing) on (a, b).

It is clear that if f is twice differentiable on (a, b) and $f''(x) > 0$ for all $x \in (a, b)$ then f is convex. A similar result also holds for concavity.

Examples : The function $f(x) = x^2$ on any open interval (in fact on all of $\mathbb{R}$) and the function $f(x) = sinx$ on $(\pi, 2\pi)$ are convex functions. The function $f(x) = -x^2$ on any open interval and the function $f(x) = sinx$ on $(0, \pi)$ are concave functions.

Definition : Let $f : (a, b) \to \mathbb{R}$ be continuous at a point $c \in (a, b)$. The point c is said to be a point of inflection if there exists a $\delta > 0$ such that either

$$f \text{ is convex on } (c-\delta, c) \text{ and } f \text{ is concave on } (c, c+\delta)$$

or

$$f \text{ is concave on } (c-\delta, c) \text{ and } f \text{ is convex on } (c, c+\delta).$$

It is clear that if $f''(x) > 0 \;\; \forall \;\; x \in (c - \delta, c)$ and $f''(x) < 0 \; \forall \;\; x \in (c, c + \delta)$ for some δ or $f''(x) < 0 \;\; \forall \;\; x \in (c - \delta, c)$ and $f''(x) > 0 \;\; \forall \;\; x \in (c, c + \delta)$ then c is a point of inflection.

Necessary Condition for Point of Inflection

Theorem 9.2: *Let $c \in (a, b)$ and $f''(c)$ exist. If f has a point of inflection at c then $f''(c) = 0$.*

Proof (*): Assume that f' is strictly increasing on $(c - \delta, c)$ and is strictly decreasing on $(c, c + \delta)$ for some $\delta > 0$. Since $f''(c)$ exists,

$$f''(c) = \lim_{x \to c^-} \frac{f'(x) - f'(c)}{x - c} \geq 0.$$

Similarly $f''(c) = \lim_{x \to c^+} \frac{f'(x)-f'(c)}{x-c} \leq 0$. Therefore $f''(0) = 0$. □

Remark : It is possible that $f''(c) = 0$ at a point but c is not a point of inflection. For example, $f(x) = x^4$ and $c = 0$. It is also possible that $f''(c)$ may not exist but c could be a point of inflection. For example $f(x) = x^{1/3}$ and $c = 0$.

Sufficient Condition for Point of Inflection

Theorem 9.3: *Let $c \in (a, b)$. If $f''(c) = 0$ and $f'''(c) \neq 0$ then c is a point of inflection.*

Proof(*): The proof is similar to the proof of Corollary 9.2 and it is left as an exercise. □

Remark : It is possible that c is a point of inflection of f and $f'''(c) = 0$. For example, consider $f(x) = x^5$ and $c = 0$.

Example : Let $f(x) = \frac{x^2-4}{x-1} = x + 1 - \frac{3}{x-1}$. Since $f'(x) > 0 \; \forall \; x \neq 1$, the function is increasing on $(-\infty, 1)$ and $(1, \infty)$ and there is no local maximum and no local minimum. Since $f''(x) > 0 \; \forall \; x < 1$ the function is convex on $(-\infty, 1)$ and since $f''(x) < 0 \; \forall \; x > 1$ the function is concave on $(1, \infty)$. There is no point of inflection as $f''(x) \neq 0$ for all $x \neq 1$ and f is not defined at $x = 1$.

Lecture 10 : Taylor's Theorem

In the last few lectures we discussed the mean value theorem (which basically relates a function and its derivative) and its applications. We will now discuss a result called Taylor's Theorem which relates a function, its derivative and its higher derivatives. We will see that Taylor's Theorem is an extension of the mean value theorem. Though Taylor's Theorem has applications in numerical methods, inequalities and local maxima and minima, it basically deals with approximation of functions by polynomials. To understand this type of approximation let us start with the linear approximation or tangent line approximation.

Linear Approximation : Let f be a function, differentiable at $x_0 \in \mathbb{R}$. Then the linear polynomial

$$P_1(x) = f(x_0) + f'(x_0)(x - x_0)$$

is the natural linear approximation to $f(x)$ near x_0. Geometrically, this is clear because we approximate the curve near $(x_0, f(x_0))$ by the tangent line at $(x_0, f(x_0))$. The following result provides an estimation of the size of the error $E_1(x) = f(x) - P_1(x)$.

Theorem 10.1: (Extended Mean Value Theorem) *If f and f' are continuous on $[a, b]$ and f' is differentiable on (a, b) then there exists $c \in (a, b)$ such that*

$$f(b) = f(a) + f'(a)(b - a) + \frac{f''(c)}{2}(b - a)^2.$$

Proof (*): This result is a particular case of Taylor's Theorem whose proof is given below.

If we take $b = x$ and $a = x_0$ in the previous result, we obtain that

$$\mid E_1(x) \mid = \mid f(x) - P_1(x) \mid \leq \frac{M}{2}(x - x_0)^2$$

where $M = sup\{\mid f''(t) \mid : t \in [x_0, x]\}$. The above estimate gives an idea "how good the approximation is i.e., how fast the error $E_1(x)$ goes to 0 as $x \to x_0$".

Naturally, one asks the question: Can we get a better estimate for the error if we use approximation by higher order polynomials. The answer is yes and this is what Taylor's theorem talks about.

There might be several ways to approximate a given function by a polynomial of degree ≥ 2, however, Taylor's theorem deals with the polynomial which agrees with f and some of its derivatives at a given point x_0 as $P_1(x)$ does in case of the linear approximation.

The polynomial

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + ... + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

has the property that $P_n(x_0) = f(x_0)$ and $P^{(k)}(x_0) = f^{(k)}(x_0)$ for all $k = 1, 2, .., n$ where $f^{(k)}(x_0)$ denotes the k th derivative of f at x_0. This polynomial is called Taylor's polynomial of degree n (with respect to f and x_0).

The following theorem called Taylor's Theorem provides an estimate for the error function $E_n(x) = f(x) - P_n(x)$.

Theorem 10.2: *Let $f : [a, b] \to \mathbb{R}, f, f', f'', ..., f^{(n-1)}$ be continuous on $[a, b]$ and suppose $f^{(n)}$ exists on (a, b). Then there exists $c \in (a, b)$ such that*

$$f(b) = f(a) + f'(a)(b - a) + \frac{f''(a)}{2!}(b - a)^2 + ... + \frac{f^{(n-1)}(a)}{(n-1)!}(b - a)^{n-1} + \frac{f^{(n)}(c)}{n!}(b - a)^n.$$

Proof (*): Define

$$F(x) = f(b) - f(x) - f'(x)(b-x) - \frac{f''(x)}{2!}(b-x)^2 - ... - \frac{f^{(n-1)}(x)}{(n-1)!}(b-x)^{n-1}.$$

We will show that $F(a) = \frac{(b-a)^n}{n!} f^{(n)}(c)$ for some $c \in (a,b)$, which will prove the theorem. Note that

$$F'(x) = -\frac{f^{(n)}(x)}{(n-1)!}(b-x)^{n-1}. \tag{1}$$

Define $g(x) = F(x) - (\frac{b-x}{b-a})^n F(a)$. It is easy to check that $g(a) = g(b) = 0$ and hence by Rolle's theorem there exists some $c \in (a,b)$ such that

$$g'(c) = F'(c) + \frac{n(b-c)^{n-1}}{(b-a)^n} F(a) = 0. \tag{2}$$

From (1) and (2) we obtain that $\frac{f^{(n)}(c)}{(n-1)!}(b-c)^{n-1} = \frac{n(b-c)^{n-1}}{(b-a)n} F(a)$. This implies that $F(a) = \frac{(b-a)^n}{n!} f^{(n)}(c)$. This proves the theorem. $\square$

Let us see some applications.

Problem 1 : Show that $1 - \frac{1}{2}x^2 \leq cosx$ for all $x \in \mathbb{R}$.

Solution : Take $f(x) = cosx$ and $x_0 = 0$ in Taylor's Theorem. Then there exists c between 0 and x such that

$$cosx = 1 - \frac{1}{2}x^2 + \frac{sinc}{6}x^3.$$

Verify that the term $\frac{sinc}{6}x^3 \geq 0$ when $\mid x \mid \leq \pi$. If $\mid x \mid \geq \pi$ then $1 - \frac{1}{2}x^2 < -3 \ \leq cosx$. Therefore the inequality holds for all $x \in \mathbb{R}$.

Problem 2 : Let $x_0 \in (a,b)$ and $n \geq 2$. Suppose $f', f'', .., f^{(n)}$ are continuous on (a,b) and $f'(x_0) = .. = f^{(n-1)}(x_0) = 0$. Then, if n is even and $f^{(n)}(x_0) > 0$, then f has a local minimum at x_0. Similarly, if n is even and $f^{(n)}(x_0) < 0$, then f has a local maximum at x_0.

Solution : By Taylor's theorem, for $x \in (a,b)$ there exists a c between x and x_0 such that

$$f(x) = f(x_0) + \frac{f^{(n)}(c)}{n!}(x - x_0)^n. \tag{3}$$

Let $f^{(n)}(x_0) > 0$ and n is even. Then by the continuity of $f^{(n)}$ there exists a neighborhood U of x_0 such that $f^{(n)}(x) > 0$ for all $x \in U$. This implies that $\frac{f^{(n)}(c)}{n!}(x - x_0)^n \geq 0$ whenever $c \in U$. Hence by equation (3), $f(x) \geq f(x_0)$ for all $x \in U$ which implies that x_0 is a local minimum.

Problem 3 : *Using Taylor's theorem, for any $k \in \mathbb{N}$ and for all $x > 0$, show that*

$$x - \frac{1}{2}x^2 + \cdots + \frac{1}{2k}x^{2k} \ < \ \log(1+x) \ < \ x - \frac{1}{2}x^2 + \cdots + \frac{1}{2k+1}x^{2k+1}.$$

Solution : By Taylor's theorem, there exists $c \in (0,x)$ s.t.

$$\log(1+x) = x - \frac{1}{2}x^2 + ... + \frac{(-1)^{n-1}}{n}x^n + \frac{(-1)^n}{n+1}\frac{x^{n+1}}{(1+c)^{n+1}}.$$

Note that, for any $x > 0$, $\frac{(-1)^n}{n+1}\frac{x^{n+1}}{(1+c)^{n+1}} > 0$ if $n = 2k$ and < 0 if $n = 2k+1$.

Lectures 11 - 13 : Infinite Series, Convergence tests, Leibniz's theorem

Series : Let (a_n) be a sequence of real numbers. Then an expression of the form $a_1 + a_2 + a_3 + ...$ denoted by $\sum_{n=1}^{\infty} a_n$, is called a series.

Examples : 1. $1 + \frac{1}{2} + \frac{1}{3} + ...$ or $\sum_{n=1}^{\infty} \frac{1}{n}$ 2. $1 + \frac{1}{4} + \frac{1}{9} + ...$ or $\sum_{n=1}^{\infty} \frac{1}{n^2}$

Partial sums : $S_n = a_1 + a_2 + a_3 + ... + a_n$ is called the nth partial sum of the series $\sum_{n=1}^{\infty} a_n$,

Convergence or Divergence of $\sum_{n=1}^{\infty} a_n$

If $S_n \to S$ for some S then we say that the series $\sum_{n=1}^{\infty} a_n$ converges to S. If (S_n) does not converge then we say that the series $\sum_{n=1}^{\infty} a_n$ diverges.

Examples :

1. $\sum_{n=1}^{\infty} log(\frac{n+1}{n})$ diverges because $S_n = log(n+1)$.
2. $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$ converges because $S_n = 1 - \frac{1}{n+1} \to 1$.
3. If $0 < x < 1$, then the geometric series $\sum_{n=0}^{\infty} x^n$ converges to $\frac{1}{1-x}$ because $S_n = \frac{1-x^{n+1}}{1-x}$.

Necessary condition for convergence

Theorem 1 : *If $\sum_{n=1}^{\infty} a_n$ converges then $a_n \to 0$.*

Proof : $S_{n+1} - S_n = a_{n+1} \to S - S = 0$. □

The condition given in the above result is necessary but not sufficient i.e., it is possible that $a_n \to 0$ and $\sum_{n=1}^{\infty} a_n$ diverges.

Examples :

1. If $\mid x \mid \geq 1$, then $\sum_{n=1}^{\infty} x^n$ diverges because $a_n \nrightarrow 0$.
2. $\sum_{n=1}^{\infty} sinn$ diverges because $a_n \nrightarrow 0$.
3. $\sum_{n=1}^{\infty} log(\frac{n+1}{n})$ diverges, however, $log(\frac{n+1}{n}) \to 0$.

Necessary and sufficient condition for convergence

Theorem 2: *Suppose $a_n \geq 0 \ \forall \ n$. Then $\sum_{n=1}^{\infty} a_n$ converges if and only if (S_n) is bounded above.*

Proof : Note that under the hypothesis, (S_n) is an increasing sequence. □

Example : The Harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges because

$$S_{2^k} \geq 1 + \frac{1}{2} + 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{8} + ... + 2^{k-1} \cdot \frac{1}{2^k} = 1 + \frac{k}{2}$$

for all k.

Theorem 3: *If $\sum_{n=1}^{\infty} \mid a_n \mid$ converges then $\sum_{n=1}^{\infty} a_n$ converges.*

Proof : Since $\sum_{n=1}^{\infty} \mid a_n \mid$ converges the sequence of partial sums of $\sum_{n=1}^{\infty} \mid a_n \mid$ satisfies the Cauchy criterion. Therefore, the sequence of partial sums of $\sum_{n=1}^{\infty} a_n$ satisfies the Cauchy criterion. □

Remark : Note that $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=p}^{\infty} a_n$ converges for any $p \geq 1$.

Tests for Convergence

Let us determine the convergence or the divergence of a series by comparing it to one whose behavior is already known.

Theorem 4 : (**Comparison test**) *Suppose* $0 \leq a_n \leq b_n$ *for* $n \geq k$ *for some* k*. Then*

(1) The convergence of $\sum_{n=1}^{\infty} b_n$ *implies the convergence of* $\sum_{n=1}^{\infty} a_n$*.*

(2) The divergence of $\sum_{n=1}^{\infty} a_n$ *implies the divergence of* $\sum_{n=1}^{\infty} b_n$*.*

Proof : (1) Note that the sequence of partial sums of $\sum_{n=1}^{\infty} a_n$ is bounded. Apply Theorem 2.

(2) This statement is the contrapositive of (1). $\square$

Examples:

1. $\sum_{n=1}^{\infty} \frac{1}{(n+1)^2}$ converges because $\frac{1}{(n+1)(n+1)} \leqslant \frac{1}{n(n+1)}$. This implies that $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.
2. $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ diverges because $\frac{1}{n} \leqslant \frac{1}{\sqrt{n}}$.
3. $\sum_{n=1}^{\infty} \frac{1}{n!}$ converges because $n^2 < n!$ for $n \geqslant 4$.

Problem 1 : *Let* $a_n \geq 0$*. Then show that both the series* $\sum\limits_{n\geq 1} a_n$ *and* $\sum\limits_{n\geq 1} \frac{a_n}{a_n+1}$ *converge or diverge together.*

Solution : Suppose $\sum\limits_{n\geq 1} a_n$ converges. Since $0 \leq \frac{a_n}{1+a_n} \leq a_n$ by comparison test $\sum\limits_{n\geq 1} \frac{a_n}{1+a_n}$ converges.

Suppose $\sum\limits_{n\geq 1} \frac{a_n}{1+a_n}$ converges. By the Theorem 1, $\frac{a_n}{1+a_n} \to 0$. Hence $a_n \to 0$ and therefore $1 \leq 1 + a_n < 2$ eventually. Hence $0 \leq \frac{1}{2}a_n \leq \frac{a_n}{1+a_n}$. Apply the comparison test.

Theorem 5 : (**Limit Comparison Test**) *Suppose* $a_n, b_n \geq 0$ *eventually. Suppose* $\frac{a_n}{b_n} \to L$*.*

1. *If* $L \in \mathbb{R}, L > 0$*, then both* $\sum_{n=1}^{\infty} b_n$ *and* $\sum_{n=1}^{\infty} a_n$ *converge or diverge together.*
2. *If* $L \in \mathbb{R}, L = 0$*, and* $\sum_{n=1}^{\infty} b_n$ *converges then* $\sum_{n=1}^{\infty} a_n$ *converges.*
3. *If* $L = \infty$ *and* $\sum_{n=1}^{\infty} b_n$ *diverges then* $\sum_{n=1}^{\infty} a_n$ *diverges.*

Proof: 1. Since $L > 0$, choose $\epsilon > 0$, such that $L - \epsilon > 0$. There exists n_0 such that $0 \leq L - \epsilon < \frac{a_n}{b_n} < L - \epsilon$. Use the comparison test.

2. For each $\epsilon > 0$, there exists n_0 such that $0 < \frac{a_n}{b_n} < \epsilon$, $\forall\ n > n_0$. Use the comparison test.

3. Given $\alpha > 0$, there exists n_0 such that $\frac{a_n}{b_n} > \alpha\ \forall\ n > n_0$. Use the comparison test. $\square$

Examples :

1. $\sum_{n=1}^{\infty}(1 - nsin\frac{1}{n})$ converges. Take $b_n = \frac{1}{n^2}$ in the previous result.
2. $\sum_{n=1}^{\infty} \frac{1}{n}log(1 + \frac{1}{n})$ converges. Take $b_n = \frac{1}{n^2}$ in the previous result.

Theorem 6 (**Cauchy Test or Cauchy condensation test**) *If* $a_n \geq 0$ *and* $a_{n+1} \leq a_n\ \forall\ n$*, then* $\sum_{n=1}^{\infty} a_n$ *converges if and only if* $\sum_{k=0}^{\infty} 2^k a_{2^k}$ *converges.*

Proof : Let $S_n = a_1 + a_2 + ... + a_n$ and $T_k = a_1 + 2a_2 + ... + 2^k a_{2^k}$.

Suppose (T_k) converges. For a fixed n, choose k such that $2^k \geq n$. Then

$$\begin{aligned} S_n &= a_1 + a_2 + ... + a_n \\ &\leq a_1 + (a_2 + a_3) + ... + (a_{2^k} + ... + a_{2^{k+1}-1}) \\ &\leq a_1 + 2a_2 + ... + 2^k a_{2^k} \\ &= T_k. \end{aligned}$$

This shows that (S_n) is bounded above; hence (S_n) converges.

Suppose (S_n) converges. For a fixed k, choose n such that $n \geq 2^k$. Then

$$\begin{aligned} S_n &= a_1 + a_2 + ... + a_n \\ &\geq a_1 + a_2 + (a_3 + a_4) ..... + (a_{2^{k-1}+1} + ... + a_{2^k}) \\ &\geq \tfrac{1}{2} a_1 + a_2 + 2a_4 + ... + 2^{k-1} a_{2^k} \\ &= \tfrac{1}{2} T_k. \end{aligned}$$

This shows that (T_k) is bounded above; hence (T_k) converges. □

Examples:

1. $\sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if $p > 1$ and diverges if $p \leq 1$.
2. $\sum_{n=1}^{\infty} \frac{1}{n(logn)^p}$ converges if $p > 1$ and diverges if $p \leq 1$.

Problem 2 : Let $a_n \geq 0, a_{n+1} \leq a_n \ \forall \ n$ and suppose $\sum a_n$ converges. Show that $na_n \to 0$ as $n \to \infty$.

Solution : By Cauchy condensation test $\sum_{k=0}^{\infty} 2^k a_{2^k}$ converges. Therefore $2^k a_{2^k} \to 0$ and hence $2^{k+1} a_{2^k} \to 0$ as $k \to \infty$. Let $2^k \leq n \leq 2^{k+1}$. Then $na_n \leq na_{2^k} \leq 2^{k+1} a_{2^k} \to 0$. This implies that $na_n \to 0$ as $n \to \infty$.

Theorem 7 (**Ratio test**) *Consider the series* $\sum_{n=1}^{\infty} a_n, \ a_n \neq 0 \ \forall \ n$.

1. *If* $\mid \frac{a_{n+1}}{a_n} \mid \leq q$ *eventually for some* $0 < q < 1$, *then* $\sum_{n=1}^{\infty} \mid a_n \mid$ *converges.*
2. *If* $\mid \frac{a_{n+1}}{a_n} \mid \geq 1$ *eventually then* $\sum_{n=1}^{\infty} a_n$ *diverges.*

Proof: 1. Note that for some N, $\mid a_{n+1} \mid \leq q \mid a_n \mid \ \ \forall \ n \ \geq N$. Therefore, $\mid a_{N+p} \mid \leq q^p \mid a_N \mid$ $\forall \ p > 0$. Apply the comparison test.

2. In this case $\mid a_n \mid \nrightarrow 0$.

Corollary 1: *Suppose* $a_n \neq 0 \ \forall \ n$, *and* $\mid \frac{a_{n+1}}{a_n} \mid \to L$ *for some* L.

1. *If* $L < 1$ *then* $\sum_{n=1}^{\infty} \mid a_n \mid$ *converges.*
2. *If* $L > 1$ *then* $\sum_{n=1}^{\infty} a_n$ *diverges.*
3. *If* $L = 1$ *we cannot make any conclusion.*

Proof :

1. Note that $\mid \frac{a_{n+1}}{a_n} \mid < L + \frac{(1-L)}{2}$ eventually. Apply the previous theorem.

2. Note that $| \frac{a_{n+1}}{a_n} | > L - \frac{(L-1)}{2}$ eventually. Apply the previous theorem.

Examples :

1. $\sum_{n=1}^{\infty} \frac{1}{n!}$ converges because $\frac{a_{n+1}}{a_n} \to 0$.

2. $\sum_{n=1}^{\infty} \frac{n^n}{n!}$ diverges because $\frac{a_{n+1}}{a_n} = (1+\frac{1}{n})^n \to e > 1$.

3. $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges and $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges, however, in both these cases $\frac{a_{n+1}}{a_n} \to 1$.

Theorem 8 : (**Root Test**) *If* $0 \leq a_n \leq x^n$ *or* $0 \leq {a_n}^{1/n} \leq x$ *eventually for some* $0 < x < 1$ *then* $\sum_{n=1}^{\infty} | a_n |$ *converges.*

Proof : Immediate from the comparison test. □

Corollary 2: *Suppose* $| a_n |^{1/n} \to L$ *for some* L*. Then*

1. *If* $L < 1$ *then* $\sum_{n=1}^{\infty} | a_n |$ *converges.*

2. *If* $L > 1$ *then* $\sum_{n=1}^{\infty} a_n$ *diverges.*

3. *If* $L = 1$ *we cannot make any conclusion.*

Examples :

1. $\sum_{n=2}^{\infty} \frac{1}{(logn)^n}$ converges because $a_n^{1/n} = \frac{1}{logn} \to 0$.

2. $\sum_{n=1}^{\infty} (\frac{n}{n+1})^{n^2}$ converges because $a_n^{1/n} = \frac{1}{(1+\frac{1}{n})^n} \to \frac{1}{e} < 1$.

3. $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges and $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges, however, in both these cases $a_n^{1/n} \to 1$.

Theorem 9 : (**Leibniz test**) *If* (a_n) *is decreasing and* $a_n \to 0$*, then* $\sum_{n=1}^{\infty} (-1)^{n+1} a_n$ *converges.*

Proof : Note that (S_{2n}) is increasing and bounded above by S_1. Similarly, (S_{2n+1}) is decreasing and bounded below by S_2. Therefore both converge. Since $S_{2n+1} - S_{2n} = a_{2n+1} \to 0$, both (S_{2n+1}) and (S_{2n}) converge to the same limit and therefore (S_n) converges. □

Examples : $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n}, \sum_{n=1}^{\infty} (-1)^n \frac{1}{n^2}$ and $\sum_{n=2}^{\infty} (-1)^n \frac{1}{logn}$ converge.

Problem 3: *Let* $\{a_n\}$ *be a decreasing sequence,* $a_n \geq 0$ *and* $\lim_{n\to\infty} a_n = 0$*. For each* $n \in \mathbb{N}$*, let* $b_n = \frac{a_1+a_2+...+a_n}{n}$*. Show that* $\sum_{n\geq 1} (-1)^n b_n$ *converges.*

Solution : Note that $b_{n+1} - b_n = \frac{1}{n+1}(a_1 + a_2 + ... + a_{n+1}) - \frac{1}{n}(a_1 + ... + a_n) = \frac{a_{n+1}}{n+1} - \frac{(a_1+...+a_n)}{n(n+1)}$. Since (a_n) is decreasing, $a_1 + ... + a_n \geq na_n$. Therefore, $b_{n+1} - b_n \leq \frac{a_{n+1}-a_n}{n+1} \leq 0$. Hence (b_n) is decreasing.

We now need to show that $b_n \to 0$. For a given $\epsilon > 0$, since $a_n \to 0$, there exists n_0 such that $a_n < \frac{\epsilon}{2}$ for all $n \geq n_0$.

Therefore, $|\frac{a_1+\cdots+a_n}{n}| = |\frac{a_1+\cdots+a_{n_0}}{n} + \frac{a_{n_0+1}+\cdots+a_n}{n}| \leq | \frac{a_1+\cdots+a_{n_0}}{n} | + \frac{n-n_0}{n}\frac{\epsilon}{2}$. Choose $N \geq n_0$ large enough so that $\frac{a_1+\cdots+a_{n_0}}{N} < \frac{\epsilon}{2}$. Then, for all $n \geq N$, $\frac{a_1+\cdots+a_n}{n} < \epsilon$. Hence, $b_n \to 0$. Use the Leibniz test for convergence.

Lecture 14 : Power Series, Taylor Series

Let $a_n \in \mathbb{R}$ for $n = 0, 1, 2, \dots$. The series $\sum_{n=0}^{\infty} a_n x^n$, $x \in \mathbb{R}$, is called a power series. More generally, if $c \in \mathbb{R}$, then the series $\sum_{n=0}^{\infty} a_n (x-c)^n$, $x \in \mathbb{R}$, is called a power series around c. If we take $x' = x - c$ then the power series around c reduces to the power series around 0. In this lecture we discuss the convergence of power series.

Examples : 1. Consider the power series $\sum_{n=0}^{\infty} \frac{1}{n!} x^n$. Let us apply the ratio test and find the set of points in $\mathbb{R}$ on which the series converges. For any $x \in \mathbb{R}$, $\frac{|a_{n+1}x^{n+1}|}{|a_n x^n|} = \frac{|x|}{n+1} \to 0$ as $n \to \infty$. Therefore the series $\sum_{n=0}^{\infty} \frac{1}{n!} x^n$ converges absolutely for all $x \in \mathbb{R}$.

2. We know that the geometric series $\sum_{n=0}^{\infty} x^n$ converges only in $(-1, 1)$. Using the ratio test we can show that the series $\sum_{n=0}^{\infty} n! x^n$ converges only at $x = 0$.

The following result gives an idea about the set on which a power series converges.

Theorem 1: *Suppose $\sum_{n=0}^{\infty} a_n x^n$ converges for some x_0 and diverges for some x_1. Then*

(i) $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely for all x such that $|\, x \,| < |\, x_0 \,|$,

(ii) $\sum_{n=0}^{\infty} a_n x^n$ diverges for all x such that $|\, x \,| > |\, x_1 \,|$.

Proof (*): (i). Suppose $x_0 \neq 0$, $\sum_{n=0}^{\infty} a_n x_0^n$ converges and $|\, x \,| < |\, x_0 \,|$. Since $\sum_{n=0}^{\infty} a_n x_0^n$ converges, there exists $M \in \mathbb{R}$ such that $|\, a_n x_0^n \,| \leq M$ for all $n \in \mathbb{N}$. Therefore,

$$|\, a_n x^n \,| = |\, a_n x_0^n \,||\, \frac{x}{x_0} \,|^n \leq M \,|\, \frac{x}{x_0} \,|^n \quad \text{for all} \quad n \in \mathbb{N}.$$

Since $|\, \frac{x}{x_0} \,| < 1$, by comparison test, the series $\sum_{n=0}^{\infty} a_n x^n$ converges absolutely.

(ii) Let $x \in \mathbb{R}$ and $|\, x \,| > |\, x_1 \,|$. Suppose $\sum_{n=0}^{\infty} a_n x^n$ converges. Then by (i), the series $\sum_{n=0}^{\infty} a_n x_1^n$ converges absolutely which is a contradiction. $\square$

From the above theorem, we can conclude that a power series $\sum_{n=0}^{\infty} a_n x^n$ is either converges for all $x \in \mathbb{R}$ or only at 0 or there is a unique $r, r > 0$ such that the series is absolutely converges for all x such that $|\, x \,| < r$ and diverges for all x such that $|\, x \,| > r$. This r is called the radius of convergence. In case the power series converges for all $x \in \mathbb{R}$ (resp., only at 0) then the radius of convergence of the series is ∞ (resp., 0).

If we define $S = \{x \in \mathbb{R} : \sum_{n=0}^{\infty} a_n x^n \text{is convergent}\}$, then the possibilities for S are

$$\{0\}, \qquad \mathbb{R}, \qquad (-r, r), \qquad [-r, r), \qquad (-r, r] \text{ and } [-r, r] \qquad \text{for some } r > 0.$$

Examples : 1. We have already seen above that the power series $\sum_{n=0}^{\infty} \frac{1}{n!} x^n$ converges for all $x \in \mathbb{R}$ and hence the radius of convergence is ∞. Similarly the radius of convergence of $\sum_{n=0}^{\infty} n! x^n$ (resp., $\sum_{n=0}^{\infty} x^n$) is 0 (resp., 1).

2. Consider the power series $\sum_{n=0}^{\infty} \frac{x^n}{n}$. Let us apply the ratio test to find the radius of convergence. For $x \in \mathbb{R}$ we have

$$\frac{|\, a_{n+1} x^{n+1} \,|}{|\, a_n x^n \,|} = |\, \frac{x^{n+1} n}{(n+1) x^n} \,| = |\, \frac{n}{n+1} x \,| \to |\, x \,| \quad \text{as} \quad n \to \infty.$$

It is clear that for $|\, x \,| < 1$ the series converges absolutely and diverges for $|\, x \,| > 1$. Therefore, the radius of convergence is 1 and the set $S = [-1, 1)$ which follows from the Leibniz test.

To find the radius of convergence of a power series or the set S, we use either the ratio test (as we did above) or root test. To find the sum of a convergent power series or for that matter sum of any convergent series is not easy. We will see sum of some particular type of power series called Taylor series.

Taylor Series : In one of the previous lectures we defined the nth degree Taylor polynomial $P_n(x)$ (w. r. to f and c), where

$$P_n(x) = f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + ... + \frac{f^{(n)}(c)}{n!}(x-c)^n.$$

The power series

$$f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + ... + \frac{f^{(n)}(c)}{n!}(x-c)^n + ... \qquad (\text{or write} \sum_{n=0}^{\infty} \frac{f^{(n)}(c)}{n!}(x-c)^n)$$

is called the Taylor series of f around c. If $c = 0$, then the Taylor series of f around c is called Macluarin series.

If f is infinite times differentiable at c then the corresponding Taylor series is defined. Moreover, $P_n(x)$ is the nth partial sum of the Taylor series. We will see in the following examples that the Taylor series may not converge for all $x \in \mathbb{R}$ and even if it converges for some x, it need not converge to $f(x)$.

Examples : 1. If we consider the function $f : \mathbb{R}\backslash\{-1, 1\} \to \mathbb{R}$ given by $f(x) = 1/(1-x)$, then the Macluarin series is the geometric series $\sum_{n=0}^{\infty} x^n$ which converges on $(-1, 1)$.

2. Define $f(x) = e^{-1/x^2}$ for $x \neq 0$ and $f(0) = 0$. Using L'Hospital rule, we can show that $f^{(k)}(0) = 0$ for all $k = 1, 2, ...$ Therefore the Macluarin series of f (for any $x \in \mathbb{R}$) is identically zero and it does not converge to $f(x)$ at any $x \neq 0$.

Taylor's theorem helps in showing the convergence of a Taylor series of f to $f(x)$ in the following way. Taylor's theorem says that there exists c_0 between x and c such that

$$E_n(x) \;=\; f(x) - P_n(x) \;=\; \frac{f^{(n+1)}(c_0)}{(n+1)!}(x-c)^{n+1}$$

It is clear from the above expression that if $E_n(x) \to 0$, then the Taylor series of f converges to $f(x)$ (as $P_n(x)$ is the nth partial sum of the Taylor series.)

Examples: Let $f(x) = sinx, x \in \mathbb{R}$. Then $\mid f^{(n)}(x) \mid \leq 1$ for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$. In this case, the Macluarin series of f converges to $f(x)$ for all $x \in \mathbb{R}$ because $E_n(x) \to 0$. (One can use the ratio test for sequence to show that $E_n(x) \to 0$). So, we can expand the sin function in the series form on whole of $\mathbb{R}$ and we write $\sin x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}, \quad x \in \mathbb{R}$.

Similarly we can show that $\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}, \; x \in \mathbb{R}$.

Problem : *Show that* $e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n, \; x > 0$.

Solution : Let $f(x) = e^x$. Fix $x > 0$. By Taylor,s Theorem there exists $c_n \in (0, x)$ such that

$$\mid E_n(x) \mid = \mid f(x) - (1 + \frac{x}{1!} + \frac{x^2}{2!} + ... + \frac{x^n}{n!}) \mid = \mid \frac{f^{n+1}(c_n)}{(n+1)!} x^{n+1} \mid = \frac{e^{c_n}}{(n+1)!} x^{n+1} \;\leq\; \frac{e^x}{(n+1)!} x^{n+1}.$$

Let $a_n = \frac{e^x}{(n+1)!} x^{n+1}$, then $\frac{a_{n+1}}{a_n} = \frac{x}{n+1} \to 0$. This implies that $a_n \to 0$ and hence $E_n(x) \to 0$.

Lecture 15-16 : Riemann Integration

Integration is concerned with the problem of finding the area of a region under a curve.

Let us start with a simple problem : *Find the area A of the region enclosed by a circle of radius r.* For an arbitrary n, consider the n equal inscribed and superscibed triangles as shown in Figure 1.

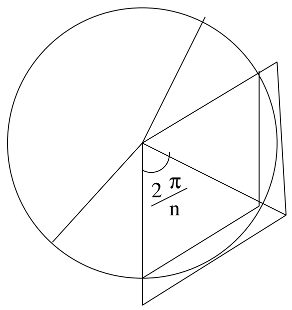


Figure 1

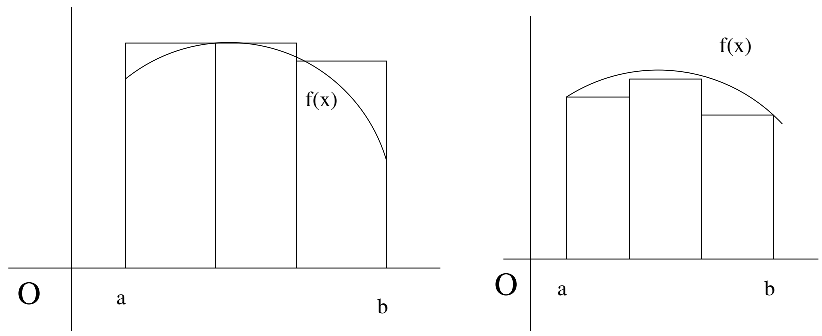


Figure 2

Since A is between the total areas of the inscribed and superscribed triangles, we have

$$nr^2 sin(\pi/n)cos(\pi/n) \leq A \leq nr^2 tan(\pi/n).$$

By sandwich theorem, $A = \pi r^2$. **We will use this idea to define and evaluate the area of the region under a graph of a function.**

Suppose f is a non-negative function defined on the interval $[a, b]$. We first subdivide the interval into a finite number of subintervals. Then we squeeze the area of the region under the graph of f between the areas of the inscribed and superscribed rectangles constructed over the subintervals as shown in Figure 2. **If** the total areas of the inscribed and superscribed rectangles converge to the same limit as we make the partition of $[a, b]$ finer and finer then the area of the region under the graph of f can be defined as this limit and f is said to be integrable.

Let us define whatever has been explained above formally.

The Riemann Integral

Let $[a, b]$ be a given interval. A *partition* P of $[a, b]$ is a finite set of points $x_0, x_1, x_2, \ldots, x_n$ such that $a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b$. We write $P = \{x_0, x_1, x_2, \ldots, x_n\}$.

If P is a partition of $[a, b]$ we write $\Delta x_i = x_i - x_{i-1}$ for $1 \leq i \leq n$. Let f be a bounded real valued function on $[a, b]$. For a partition P of $[a, b]$, we define

$$M_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\} \ \ and \ \ m_i = \inf\{f(x) : x_{i-1} \leq x \leq x_i\}.$$

$$U(P, f) = \sum_1^n M_i \Delta x_i \ \ and \ \ L(P, f) = \sum_1^n m_i \Delta x_i.$$

The numbers $U(P, f)$ and $L(P, f)$ are called *upper and lower Riemann sums* for the partition P (see Figure 2).

Since f is bounded, there exist real numbers m and M such that $m \leq f(x) \leq M$, for all $x \in [a, b]$. Thus for every partition P,

$$m(b - a) \leq L(P, f) \leq U(P, f) \leq M(b - a).$$

We define

$$\overline{\int_a^b} f dx = \inf U(P, f) \tag{1}$$

and

$$\underline{\int_a^b} f dx = \sup L(P, f). \tag{2}$$

(1) and (2) are called *upper and lower Riemann integrals* of f over $[a, b]$ respectively.

If the upper and lower integrals are equal, we say that f *is Riemann integrable* or *integrable.* In this case the common value of (1) and (2) is called the Riemann integral of f and is denoted by $\int_a^b f dx$ or $\int_a^b f(x)dx$.

Examples : 1. Consider the function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(\frac{1}{2}) = 1 \text{ and } f(x) = 0 \text{ for all } x \in [0, 1] \backslash \{\frac{1}{2}\}.$$

Then f is integrable. We show this using the definition as follows. For any partition P of $[0, 1]$, $L(P, f)$ is always 0 and hence the lower integral is 0. Let us evaluate the upper integral. Let $P = \{x_1, x_2, .., x_n\}$ be any partition of $[0, 1]$ and $\frac{1}{2} \in [x_i, x_{i+1}]$ for some i. Then $U(P, f) \leq 2max\Delta x_j$. Since we can always choose a partition P such that $max\Delta x_j$ is as small as possible, the upper integral, which is the infimum of $U(P, f)'s$, is 0. Hence, f is integrable and $\int_0^1 f(x)dx = 0$.

2. Not every bounded function is integrable. For example the function

$$f(x) = 1 \;\; if \; x \; is \; rational \; and \; 0 \; otherwise$$

is not integrable over any interval $[a, b]$ (Check this).

In general, determining whether a bounded function on $[a, b]$ is integrable, using the definition, is difficult. For the purpose of checking the integrability, we give a criterion for integrability, called Riemann criterion, which is analogous to the Cauchy criterion for the convergence of a sequence.

Let us define some concepts and results before presenting the criterion. Throughout, we will assume that f is a bounded real function on $[a, b]$.

Definition: A partition P_2 of $[a, b]$ is said to be finer than a partition P_1 if $P_2 \supset P_1$. In this case we say that P_2 is a *refinement* of P_1. Given two partition P_1 and P_2, the partition $P_1 \cup P_2 = P$ is called their common refinement.

The following theorem illustrates that refining partition increases lower terms and decreases upper terms.

Theorem 1 : *Let P_2 be a refinement of P_1 then $L(P_1, f) \leq L(P_2, f)$ and $U(P_2, f) \leq U(P_1, f)$.*

Proof (*): First we assume that P_2 contains just one more point than P_1. Let this extra point be $x^\star$. Suppose $x_{i-1} < x^\star < x_i$, where x_{i-1} and x_i are consecutive points of P_1. Let

$$\begin{aligned} w_1 &= \inf\{f(x) : \; x_{i-1} \leq x \leq x^\star\} \text{ and} \\ w_2 &= \inf\{f(x) : \; x^\star \leq x \leq x_i.\} \end{aligned}$$

Then $w_1 \geq m_i$ and $w_2 \geq m_i$ where $m_i = \inf\{f(x) : \; x_{i-1} \leq x \leq x_i\}$. Then

$$L(P_2, f) - L(P_1, f) = w_1(x^\star - x_{i-1}) + w_2(x_i - x^\star) - m_i(x_i - x_{i-1}) \geq 0.$$

If P_2 contains k more points then we repeat this process $k-$times. The other inequality is analogously proved. (Prove it). □

The geometric interpretation suggests that the lower integral is less than or equal to the upper integral. So the next result is also anticipated.

Corollary 2 : $\overline{\int_a^b} f dx \geq \underline{\int_a^b} f dx$.

Proof (*): Let P_1, P_2 be two partitions and let P be their common refinement. Then

$$L(P_1, f) \leq L(P, f) \leq U(P, f) \leq U(P_2, f).$$

Thus for any two partitions P_1 and P_2, we have $L(P_1, f) \leq U(P_2, f)$.
Fix P_2 and take sup over all P_1. Then $\underline{\int_a^b} f dx \leq U(P_2, f)$. Now take inf over all P_2 to get the desired result. □

In the following result we present the Reimann criterion (a necessary and sufficient condition for the existence of the integral of a bounded function).

Theorem 3 : (Riemann's criterion for integrability): *f is integrable on $[a, b] \Leftrightarrow$ for every $\epsilon > 0$ there exists a partition P such that*

$$U(P, f) - L(P, f) < \epsilon. \tag{3}$$

Proof (*): For any P, we have

$$L(P, f) \leq \underline{\int_a^b} f dx \leq \overline{\int_a^b} f dx \leq U(P, f).$$

Therefore (3) implies that

$$\overline{\int_a^b} f dx - \underline{\int_a^b} f dx < \epsilon, \quad \forall\ \epsilon\ >\ 0.$$

Hence $\underline{\int_a^b} f dx = \overline{\int_a^b} f dx$ i.e. f is integrable. Conversely, suppose f is integrable and $\epsilon > 0$. Then there exist partitions P_1 and P_2 such that

$$U(P_2, f) - \int_a^b f dx < \epsilon/2 \text{ and } \int_a^b f dx - L(P_1, f) < \epsilon/2$$

Let P be the common refinement of P_1 and P_2. Then $U(P, f) - L(P, f) < \epsilon$. □

The proof of the following corollary is immediate from the previous theorem.

Corollary 3 : Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Suppose (P_n) is a sequence of partitions of $[a, b]$ such that $U(P_n, f) - L(P_n, f) \to 0$, then f is integrable.

Problem : Let $f : [0,\ 1] \to \mathbb{R}$ such that $f(x) = \begin{cases} \frac{1}{n} & \text{if } x = \frac{1}{n} \\ 0 & \text{otherwise.} \end{cases}$ Show that f is integrable and find $\int_0^1 f(x) dx$.

Solution : We will use the Riemann criterion to show that f is integrable on $[0, 1]$. Let $\varepsilon > 0$ be given. We will choose a partition P such that $UP, f) - L(P, f) < \varepsilon$. Since $1/n \to 0$, there exists N such that $1/n \in [0, \varepsilon]$ for all $n > N$. So only finite number of $\frac{1}{n}'s$ lie in the interval $[\epsilon, 1]$. Cover these finite number of $\frac{1}{n}'s$ by the intervals $[x_1, x_2], [x_3, x_4], ..., [x_{m-1}, x_m]$ such that $x_i \in [\varepsilon, 1]$ for all

$i = 1, 2, .., m$ and the sum of the length of these m intervals is less than ε. Consider the partition $P = \{0, \varepsilon, x_1, x_2, ..., x_m\}$. It is clear that $U(P, f) - L(P, f) < 2\varepsilon$. Hence by the Reimann criterion the function is integrable. Since the lower integral is 0 and the function is integrable, $\int_0^1 f(x)dx = 0$.

We will apply the Riemann criterion for integrability to prove the following two existence theorems.

Theorem 4: *If f is continuous on $[a, b]$ then f is integrable.*

Proof : Let $\epsilon > 0$. Since f is uniformly continuous, choose $\delta > 0$ such that $|s - t| < \delta \Rightarrow |f(s) - f(t)| < \epsilon$ for $s, t \in [a, b]$.

Let P be a partition of $[a, b]$ such that $\Delta x_i < \delta \ \forall\, i = 1, 2, \ldots, n$. Then

$$M_i - m_i \le \epsilon \ \forall\, i = 1, 2, \ldots, n.$$

Hence

$$U(P, f) - L(P, f) = \sum_1^n (M_i - m_i)\Delta x_i \le \epsilon(b - a).$$

This implies that f is integrable. □

Theorem 5: *If f is a monotone function then f integrable.*

Proof : Suppose f is monotonically increasing (the proof is similar in the other case.)
Choose a partition P such that $\Delta x_i = \frac{b-a}{n}$. Then $M_i = f(x_i)$ and $m_i = f(x_{i-1})$. Therefore

$$\begin{aligned} U(P, f) - L(P, f) &= \frac{b-a}{n}\sum_1^n [f(x_i) - f(x_{i-1})] \\ &= \frac{b-a}{n}[f(b) - f(a)] \\ &< \epsilon \qquad \text{for large } n. \end{aligned}$$

Hence f is integrable. □

In the following problem we will see that limit and integral cannot be interchanged.

Problem : *Let* $g_n(y) = \begin{cases} \frac{ny^{n-1}}{1+y} & \text{if } 0 \le y < 1 \\ 0 & \text{if } y = 1 \end{cases}$. *Then prove that* $\lim\limits_{n\to\infty} \int\limits_0^1 g_n(y)dy = \frac{1}{2}$ *whereas* $\int\limits_0^1 \lim\limits_{n\to\infty} g_n(y)dy = 0$.

Solution : From the ratio test for sequences we can show that $\lim\limits_{n\to\infty} \frac{ny^{n-1}}{1+y} = 0$, for each $0 < y < 1$. Therefore $\int_0^1 \lim\limits_{n\to\infty} g_n(y)dy = 0$.

For the other part, use integration by parts to see that $\int\limits_0^1 \frac{ny^{n-1}}{1+y}dy = \frac{1}{2} + \int\limits_0^1 \frac{y^n}{(1+y)^2}dy$. Note that $\int\limits_0^1 \frac{y^n}{(1+y)^2}dy \le \int\limits_0^1 y^n = \frac{1}{n+1} \to 0$ as $n \to \infty$. Therefore, $\lim_{n\to\infty} \int\limits_0^1 g_n(y)dy = \frac{1}{2}$.

Lecture 17 : Fundamental Theorems of Calculus, Riemann Sum

By looking at the definitions of differentiation and integration, one may feel that these notions are totally different. Even the geometric interpretations do not give any idea that these two notions are related. In this lecture we will discuss two results, called fundamental theorems of calculus, which say that differentiation and integration are, in a sense, inverse operations.

Theorem 17.1: ***(First Fundamental Theorem of Calculus)*** Let f be integrable on $[a,b]$. For $a \leq x \leq b$, let $F(x) = \int_a^x f(t)dt$. Then F is continuous on $[a,b]$ and if f is continuous at x_0 then F is differentiable at x_0 and $F'(x_0) = f(x_0)$.

Proof (*): Suppose $M = \sup\{|f(x)| : x \in [a,b]\}$. Let $a \leq x < y \leq b$. Then

$$|F(y) - F(x)| = |\int_x^y f(t)dt| \leq M(y-x).$$

Thus $|F(x) - F(y)| \leq M|x-y|$ for $x, y \in [a,b]$. Hence F is continuous, in fact uniformly continuous.

Now suppose f is continuous at x_0. Given $\epsilon > 0$ choose $\delta > 0$ such that

$$|t - x_0| < \delta \ \Rightarrow |f(t) - f(x_0)| < \epsilon.$$

Let x be such that $0 \leq |x - x_0| < \delta$. Then

$$\left|\frac{F(x) - F(x_0)}{x - x_0} - f(x_0)\right| = \left|\frac{1}{x - x_0}\int_{x_0}^x [f(t) - f(x_0)]dt\right| < \epsilon.$$

This implies that $F'(x_0) = f(x_0)$. □

In the previous theorem, in a sense, we obtained f by differentiating integral of f when f is continuous on $[a,b]$. A function F such that $F'(x) = f(x)\ \forall\ x \in [a,b]$ is called an antiderivative of f on $[a,b]$. The existence of an antiderivative for a continuous function on $[a,b]$ follows from the first F.T.C.

If an integrable function f has an antiderivative (and if we can find it), then calculating its integral is very simple. The second F.T.C. explains this.

Theorem 17.2 : ***(Second Fundamental Theorem of Calculus)*** Let f be integrable on $[a,b]$. If there is a differentiable function F on $[a,b]$ such that $F' = f$ then $\int_a^b f(x)dx = F(b) - F(a)$.

Proof (*): Let $\epsilon > 0$. Since f is integrable we can find a partition $P := \{x_0, x_1, ..., x_n\}$ of $[a,b]$ such that $U(P,f) - L(P,f) < \epsilon$. By the mean value theorem there exists $c_i \in [x_{i-1}, x_i]$ such that $F(x_i) - F(x_{i-1}) = f(c_i)\Delta x_i$. Hence

$$\sum f(c_i)\Delta x_i = F(b) - F(a).$$

we know that

$$L(P,f) \leq \int_a^b f dx \leq U(P,f)$$

and

$$L(P,f) \leq \sum f(c_i)\Delta x_i \leq U(P,f).$$

Therefore $|F(b) - F(a) - \int_a^b f dx| < \epsilon$. This completes the proof. □

The second F.T.C. explains why the indefinite integral of F' is defined to be F.

Remark : The proof of Theorem 17.2 becomes simpler if, instead of assuming f to be integrable, we make stonger assumption that f is continuous on $[a, b]$. In fact, the proof follows from Theorem 17.1 (prove !).

Problem 1: *Let p be a fixed number and let f be a continuous function on $\mathbb{R}$ that satisfies the equation $f(x + p) = f(x)$ for every $x \in \mathbb{R}$. Show that the integral $\int_a^{a+p} f(t)dt$ has the same value for every real number a.*

Solution : Suppose $a, p > 0$. Then by the first F.T.C., we have

$$\frac{d}{da}\Big(\int_a^{a+p} f(t)dt\Big) = \frac{d}{da}\Big(\int_0^{a+p} f(t)dt - \int_0^a f(t)dt\Big) = f(a + p) - f(a) = 0.$$

Problem 2: *Let f be a continuous function on $[0, \pi/2]$ and $\int_0^{\pi/2} f(t)dt = 0$. Show that there exists a $c \in (0, \pi/2)$ such that $f(c) = 2cos2c$.*

Solution : Define F on $[0, \frac{\pi}{2}]$ such that $F(x) = \int\limits_0^x f(t)dt - \sin 2x$. Apply the first F.T.C. and Rolle's theorem.

Problem 3: *Show that $lim_{x\to 0}\frac{1}{x^3}\int\limits_0^x \frac{t^2}{1+t^4}dt = \frac{1}{3}$.*

Solution : Apply the first F.T.C. and the L'Hospital rule.

Riemann Sum

We now see an important property of integrable functions.

Definition: Let $f : [a, b] \to \mathbb{R}$ and let $P = \{x_0, x_1, ..., x_n\}$ be a partition of $[a, b]$. Let $c_k \in [x_{k-1}, x_k], k = 1, 2, ..., n$. Then a *Riemann sum* for f (corresponding to the partition P and the intermediate points c_k) is $S(P, f) = \sum_{k=1}^n f(c_k)\Delta x_k$.

The norm of P is defined by $\| P \| = max_{1\le i\le n}\Delta x_i$.

Theorem 4: *Let $f : [a, b] \to \mathbb{R}$ be integrable. Then $\lim\limits_{\|P\|\to 0} S(P, f) = \int_a^b f(x)dx$.*

We will not present the proof of this theorem, however, we will use it later when we discuss the applications of integration. This result also has some other applications. For example, we can use this to approximate the integral of f when we cannot evaluate it exactly. Moreover, this result can also be used to find limit of certain type of sequences.

Example : Let us evaluate $\lim_{n\to\infty} x_n$ where $x_n = \frac{1}{n} + \frac{1}{n+1} + ... + \frac{1}{2n-1}$ using the above theorem. Basically we have to write x_n as a Riemann sum of some function on some interval. Note that

$$x_n = \frac{1}{n}\Big(\frac{1}{1} + \frac{1}{1 + \frac{1}{n}} + \frac{1}{1 + \frac{2}{n}} + ... + \frac{1}{1 + \frac{n-1}{n}}\Big) = S(P_n, f)$$

where $f(x) = \frac{1}{x}$ for $x \in [1, 2]$ and $P_n = \{1, 1 + \frac{1}{n}, 1 + \frac{2}{n}, ..., 1 + \frac{n}{n}\}$. Note that here $c_{k+1} = \frac{1}{1+\frac{1}{k}}$. Therefore, by the previous theorem we have $\lim_{n\to\infty} x_n = \int_1^2 \frac{1}{x}dx$.

Lecture 18 : Improper integrals

We defined $\int_a^b f(t)dt$ under the conditions that f is defined and bounded on the bounded interval $[a,b]$. In this lecture, we will extend the theory of integration to bounded functions defined on unbounded intervals and also to unbounded functions defined on bounded or unbounded intervals.

Improper integral of the first kind: Suppose f is (Riemann) integrable on $[a,x]$ for all $x > a$, i.e., $\int_a^x f(t)dt$ exists for all $x > a$. If $\lim\limits_{x\to\infty} \int_a^x f(t)dt = L$ for some $L \in \mathbb{R}$, then we say that the improper integral (of the first kind) $\int_a^\infty f(t)dt$ converges to L and we write $\int_a^\infty f(t)dt = L$. Otherwise, we say that the improper integral $\int_a^\infty f(t)dt$ diverges.

Observe that the definition of convergence of improper integrals is similar to the one given for series. For example, $\int_a^x f(t)dt,\ x > a$ is analogous to the partial sum of a series.

Examples : 1. The improper integral $\int_1^\infty \frac{1}{t^2}dt$ converges, because, $\int_1^x \frac{1}{t^2}dt = 1 - \frac{1}{x} \to 1$ as $x \to \infty$. On the other hand, $\int_1^\infty \frac{1}{t}dt$ diverges because $\lim_{x\to\infty} \int_1^x \frac{1}{t}dt = \lim_{x\to\infty} \log x$. In fact, one can show that $\int_1^\infty \frac{1}{t^p}dt$ converges to $\frac{1}{p-1}$ for $p > 1$ and diverges for $p \leq 1$.

2. Consider $\int_0^\infty te^{-t^2}dt$. We will use substitution in this example. Note that

$$\int_0^x te^{-t^2}dt = \frac{1}{2}\int_0^{x^2} e^{-s}ds = \frac{1}{2}(1 - e^{-x^2}) \to \frac{1}{2}\ as\ x \to \infty.$$

3. The integral $\int_0^\infty sintdt$ diverges, because, $\int_0^x sintdt = 1 - cosx$.

We now derive some convergence tests for improper integrals. These tests are similar to those used for series. We do not present the proofs of the following three results as they are similar to the proofs of the corresponding results for series.

Theorem 17.1 : *Suppose f is integrable on $[a,x]$ for all $x > a$ where $f(t) \geq 0$ for all $t > a$. If there exists $M > 0$ such that $\int_a^x f(t)dt \leq M$ for all $x \geq a$ then $\int_a^\infty f(t)dt$ converges.*

This result is similar to the result: *If $a_n \geq 0$ for all n and the partial sum $S_n \leq M$ for all n, then $\sum a_n$ converges.* The proofs are also similar. One uses the above theorem to prove the following theorem which is analogous to the comparison test of series.

In the following two results we assume that f and g are integrable on $[a,x]$ for all $x > a$.

Theorem 17.2 : (Comparison test) *Suppose $0 \leq f(t) \leq g(t)$ for all $t > a$. If $\int_a^\infty g(t)dt$ converges, then $\int_a^\infty f(t)dt$ converges.*

Examples : 1. The improper integral $\int_1^\infty \frac{cos^2 t}{t^2}dt$ converges, because $0 \leq \frac{cos^2 t}{t^2} \leq \frac{1}{t^2}$.

2. The improper integral $\int_1^\infty \frac{2+sint}{t}dt$ diverges, because $\frac{2+sint}{t} \geq \frac{1}{t} > 0$ for all $t > 1$.

Theorem 17.3 : (Limit Comparison Test(LCT)) *Suppose $f(t) \geq 0$ and $g(t) > 0$ for all $x > a$. If $\lim\limits_{t\to\infty} \frac{f(t)}{g(t)} = c$ where $c \neq 0$, then both the integrals $\int_a^\infty f(t)dt$ and $\int_a^\infty g(t)dt$ converge or both diverge. In case $c = 0$, then convergence of $\int_a^\infty g(t)dt$ implies convergence of $\int_a^\infty f(t)dt$.*

Examples : 1. The integral $\int_1^\infty sin\frac{1}{t}dt$ diverges by LCT, because $\frac{sin\frac{1}{t}}{\frac{1}{t}} \to 1$ as $t \to \infty$.

2. For $p \in \mathbb{R}$, $\int_1^\infty e^{-t}t^p dt$ converges by LCT because $\frac{e^{-t}t^p}{t^{-2}} \to 0$ as $x \to \infty$.

So far we considered the convergence of improper integrals of only non-negative functions. We will now consider any real valued functions. The following result is anticipated.

Theorem 17.4 : *If an improper integral* $\int_a^\infty \mid f(t) \mid dt$ *converges then* $\int_a^\infty f(t)dt$ *converges i.e., every absolutely convergent improper integral is convergent.*

Proof : Suppose $\int_a^\infty \mid f(t) \mid dt$ converges and $\int_a^x f(t)dt$ exists for all $x > a$. Since $0 \leq f(x)+ \mid f(x) \mid \leq 2 \mid f(x) \mid$, by comparison test $\int_a^\infty (f(x)+ \mid f(x) \mid)dx$ converges. This implies that $\int_a^\infty (f(x)+ \mid f(x) \mid - \mid f(x) \mid)dx$ converges. □

The converse of the above theorem is not true (see Problem 2).

The following result, known as **Dirichlet test**, is very useful.

Theorem 17.5 : *Let* $f, g : [a, \infty) \to \mathbb{R}$ *be such that*

(i) f *is decreasing and* $f(t) \to 0$ *as* $t \to \infty$*,*

(ii) g *is continuous and there exists* M *such that* $\int_a^x g(t)dt \leq M$ *for all* $x > a$*.*

Then $\int_a^\infty f(t)g(t)dt$ *converges.*

We will not present the proof of the above theorem but we will use it.

Examples : Integrals $\int_\pi^\infty \frac{sint}{t}dt$ and $\int_\pi^\infty \frac{\cos t}{t}dt$ are convergent.

Improper integrals of the form $\int_{-\infty}^b f(t)dt$ are defined similarly. We say that $\int_{-\infty}^\infty f(t)dt$ is convergent if both $\int_{-\infty}^c f(t)dt$ and $\int_c^\infty f(t)dt$ are convergent for some element c in $\mathbb{R}$ and $\int_{-\infty}^\infty f(t)dt = \int_{-\infty}^c f(t)dt + \int_c^\infty f(t)dt$.

Improper integral of second kind : Suppose $\int_x^b f(t)dt$ exists for all x such that $a < x \leq b$ (the function f could be unbounded on $(a, b]$). If $\lim\limits_{x \to a^+} \int_x^b f(t)dt = M$ for some $M \in \mathbb{R}$, then we say that the improper integral (of the second kind) $\int_a^b f(t)dt$ converges to M and we write $\int_a^b f(t)dt = M$.

Example : The improper integral $\int_0^1 \frac{1}{t^p}dt$ converges for $p < 1$ and diverges for $p \geq 1$.

Comparison test and limit comparison test for improper integral of the second kind are analogous to those of the first kind. If an improper integral is a combination of both first and second kind then one defines the convergence similar to that of the improper integral of the kind $\int_{-\infty}^\infty f(t)dt$,

Problem 1: Determine the values of p for which $\int\limits_0^\infty f(x)dx$ converges where $f(x) = \frac{1-e^{-x}}{x^p}$.

Solution : Let $I_1 = \int\limits_0^1 f(x)dx$ and $I_2 = \int\limits_1^\infty f(x)dx$. We have to determine the values of p for which the integrals I_1 and I_2 converge. Now one has to see how the function $f(x)$ behaves in the respective intervals and apply the LCT. Since $\lim\limits_{x \to 0} \frac{1-e^{-x}}{x} = 1$, by LCT with $\frac{1}{x^{p-1}}$, we see that I_1 is convergent iff $p - 1 < 1$, *i.e.*,$p < 2$. Similarly, I_2 is convergent (by applying LCT with $\frac{1}{x^p}$) iff $p > 1$. Therefore $\int\limits_0^\infty f(x)dx$ converges iff $1 < p < 2$.

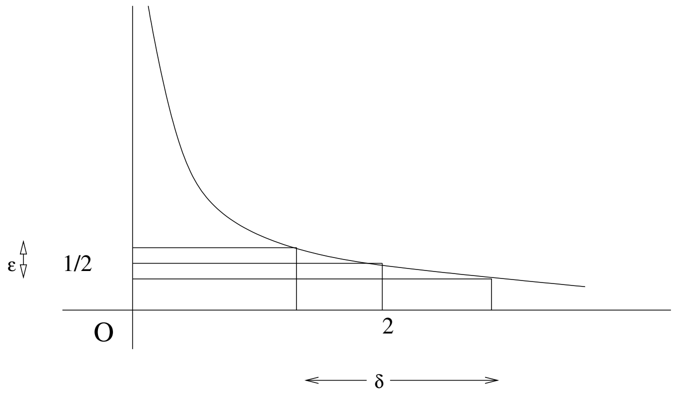
Problem 2 : Prove that $\int_1^\infty \frac{\sin x}{x^p}dx$ converges but not absolutely for $0 < p \leq 1$.

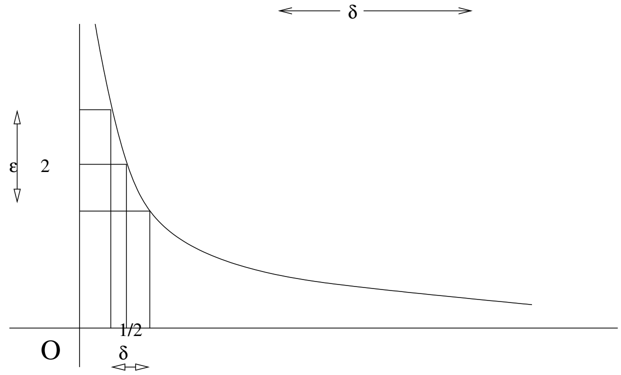
Solution : Let $0 < p \leq 1$. By Dirichlet's Test, the integral converges. We claim that $\int\limits_1^\infty \frac{|\sin x|}{x^p}dx$ does not converge. Since, $|\sin x| \geq \sin^2 x$, we see that $|\frac{\sin x}{x^p}| \geq \frac{\sin^2 x}{x^p} = \frac{1-\cos 2x}{2x^p}$. By Dirichlet's Test, $\int\limits_1^\infty \frac{\cos 2x}{2x^p}dx$ converges $\forall\ p > 0$. But $\int\limits_1^\infty \frac{1}{2x^p}$ diverges for $p \leq 1$. Hence, $\int\limits_1^\infty \mid \frac{\sin x}{x^p} \mid dx$ does not converge.

Uniform Continuity

Let us first review the notion of continuity of a function. Let $A \subset I\!R$ and $f : A \to I\!R$ be continuous. Then for each $x_0 \in A$ and for given $\varepsilon > 0$, there exists a $\delta(\varepsilon, x_0) > 0$ such that $x \epsilon A$ and $\mid x - x_0 \mid < \delta$ imply $\mid f(x) - f(x_0) \mid < \varepsilon$. We emphasize that δ depends, in general, on ϵ as well as the point x_0. Intuitively this is clear because the function f may change its values rapidly near certain points and slowly near other points.

For example consider $f(x) = 1/x$. The following two figures explain that for a given ε- neighbourhood about each of $f(2) = 1/2$ and $f(1/2) = 2$, the corresponding maximum values of δ for the points 2 and 1/2 are seen to be different.





We also see that as x_0 tends to 0, the permissible values of δ tends to 0.

Example 1: Consider the function $f : I\!R \to I\!R$ defined by $f(x) = x^2$ for all $x \in I\!R$. Suppose $\varepsilon = 2$ and $x_0 = 1$. Then $f(x) - f(x_0) = x^2 - 1$. If $\mid x - 1 \mid < 1/2$ then $1/2 < x < 3/2$ and so $-3/4 < x^2 - 1 < 5/4$. Therefore with $\varepsilon = 2$ and $x_0 = 1$, we have $\mid x - x_0 \mid < 1/2$ imply $\mid f(x) - f(x_0) \mid < 2$. So $\delta = 1/2$ works in this case.

We will now illustrate that the previous statement is not true for $x_0 = 10$. For, when $x_0 = 10$ we have $f(x) - f(x_0) = x^2 - 100$. If $x = 10 + 1/4$ then $\mid x - x_0 \mid < 1/2$ but $f(x) - f(x_0) = (10 + 1/4)^2 - 10^2 > 2$. This shows that even though f is continuous at the point 10 as well at the point 1, for $\varepsilon = 2$ the number $\delta = 1/2$ works for $x_0 = 1$ but not for $x_0 = 10$.

One may ask that for this f, corresponding to $\varepsilon = 2$, there might be some δ (possibly depending on ε) that will work for all $x \in I\!R$. We will show that the answer to this question is negative. Suppose there is a $\delta > 0$ such that for every $x \in I\!R$, we have:

$$| x - y | < \delta \text{ imply } | f(x) - f(y) | < 2.$$

Let $x \in I\!R$ and choose $y = x + \delta/2$. Since $| x - y | < \delta$, by assumption, we have

$$\begin{aligned} | f(x) - f(y) | &= | x - y || x + y | \\ &= \delta/2 \, | 2x + \delta/2 | \\ &= | \delta x + \delta^2/4 | < 2 \end{aligned}$$

This implies that $\delta x < 2$ for all $x \in I\!R^+$, the set of positive real numbers. This is clearly false.

The next example shows that it is not always the case that δ is dependent upon $x_0 \epsilon A$.

Example 2 : Let $A = \{x \in I\!R : x \geq 0\} = [0, \infty) \subset I\!R$ and $f : A \to I\!R$ be defined by

$$f(x) = \sqrt{x}.$$

It is easy to verify that for all $x, y \in A$, $| f(x) - f(y) | = | \sqrt{x} - \sqrt{y} | \leq \sqrt{| x - y |}$. Therefore for given $\varepsilon > 0$ if we choose $\delta = \varepsilon^2$. We have :

$$x, y \in A \text{ and } | x - y | < \delta \text{ imply } | f(x) - f(y) | < \varepsilon.$$

The preceding discussion motivates the following definition.

Definition: A function $f : A \to I\!R$, where $A \subset I\!R$ is said to be *uniformly continuous on* A if given $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $x, y \in A$ and $| x - y | < \delta$, we have $| f(x) - f(y) | < \varepsilon$

Clearly uniform continuity implies continuity but the converse is not always true as seen from Example 1.

In the previous definition we also emphasise that the uniform continuity of f is dependent upon the function f and on the set A. For example, we had seen in Example 1 that the function defined by $f(x) = x^2$ is not uniformly continuous on $I\!R$ or (a, ∞) for all $a \in I\!R$. Let $A = [a, b], a > 0$ and $\varepsilon > 0$. Then

$$| f(x) - f(y) | = | x^2 - y^2 | = | x - y || x + y | \leq 2b \, | x - y |$$

Hence for $\delta = \dfrac{\varepsilon}{2b}$. We have

$$x, y \in I\!R, | x - y | < \delta \Rightarrow | f(x) - f(y) | < \varepsilon$$

Therefore f is uniformly continuous on $[a, b]$.

Infact we illustrate that every continuous function on any closed bounded interval is uniformly continuous.

Let us formulate an equivalent condition to saying that f is not uniformly continuous on A.

Let $A \subset I\!R$ and let $f : A \to I\!R$. Then the following conditions are equivalent.

(i) f is not uniformly condtion on A .

(ii) There exists an $\epsilon_0 > 0$ such that for every $\delta > 0$ there are points x, y in A such that $| x - y | < \delta$ and $| f(x) - f(y) | \geq \varepsilon_0$.

(iii) There exist an $\epsilon_0 > 0$ and two sequences (x_n) and (y_n) in A such that $\lim(x_n - y_n) = 0$ and $| f(x_n) - f(y_n) | \geq \varepsilon_0$ for all $n \in N$.

Example 3: *We can apply this result to show that $g(x) := \dfrac{1}{x}$ is not uniformly continuous on $A := \{x \epsilon I\!R : x > 0\}$. For if $x_n := \dfrac{1}{n}$ and $y_n := \dfrac{1}{n+1}$, then we have $\lim(x_n - y_n) = 0$ but $| g(x_n) - g(y_n) | = 1$ for all $n \in N$.*

As an immediate consequence of the previous observation, we have the following result which provides us with a sequential criterion for uniform continuity.

Prososition 1: *A function $f : A \to I\!R$ is uniformly continuous on a set $A \subset I\!R$ if and only if whenever sequences (x_n) and (y_n) of points A are such that the sequence $(x_n - y_n)$ converges to 0, the sequence $f(x_n) - f(y_n)$ converges to 0.*

Theorem 2: *Let $a < b$ and $f : [a, b] \to I\!R$ be continuous. Then f is uniformly continuous.*

Proof: Assume the contrary that f is not uniformly continuous. Hence there exist an $\epsilon_0 > 0$ and two sequences (x_n) and (y_n) in $[a, b]$ such that $x_n - y_n \to 0$ and $| f(x_n) - f(y_n) | \geq \varepsilon_0$ for all $n \in N$. Since (x_n) is in $[a, b]$, by Theorem 2.8, there exists a subsequence (x_{ni}) of (x_n) such that $x_{ni} \to x_0 \in [a, b]$. Hence $y_{ni} \to x_0$. By continuity of f, it follows that $f(x_{ni}) \to f(x_0)$ and $f(y_{ni}) \to f(x_0)$. Therefore $| f(x_{ni}) - f(y_{ni}) | \to 0$. This contradicts the fact that $| f(x_{ni}) - f(y_{ni}) | \geq \varepsilon_0$. Therefore f is uniformly continuous.

Problems :

1. Let $A \subset I\!R$ and $f : A \to I\!R$ be uniformly continuous on A. Show that if (x_n) is a Cauchy sequence in A then $(f(x_n))$ is a Cauchy sequence in $I\!R$.

2. Using the previous problem show that the following functions are not uniformly continuous.
(i) $f(x) = \dfrac{1}{x^2}, x \in (0, 1)$

(ii) $f(x) = tanx, x \in [0, \dfrac{\pi}{2})$

Lecture 19: Area between two curves; Polar coordinates

Recall that our motivation to introduce the concept of a Riemann integral was to define (or to give a meaning to) the area of the region under the graph of a function. If $f : [a, b] \to \mathbb{R}$ be a continuous function and $f(x) \geq 0$ then the area of the region between the graph of f and the x-axis is defined to be

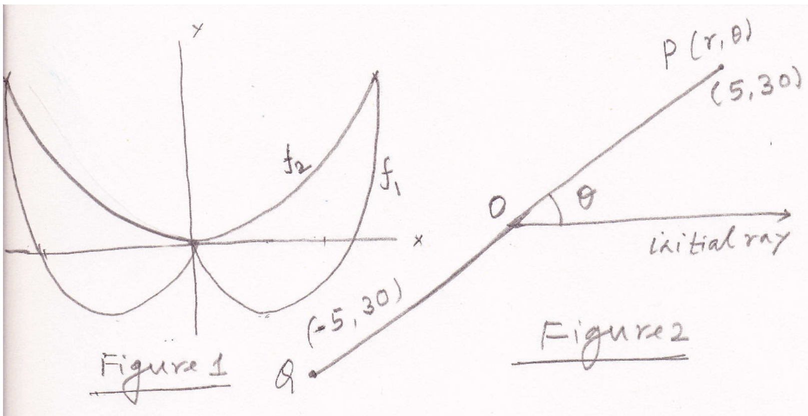
$$Area = \int_a^b f(x)dx.$$

Instead of the x-axis, we can take a graph of another continuous function $g(x)$ such that $g(x) \leq f(x)$ for all $x \in [a, b]$ and define the area of the region between the graphs to be

$$Area = \int_a^b (f(x) - g(x))\, dx.$$

Examples: 1. Let us find the area bounded by the curves: $f_1(x) = x^4 - 2x^2$ and $f_2(x) = 2x^2$. The common points of intersection of the graphs are the points satisfying : $f_1(x) = f_2(x)$ i.e., $x^4 - 2x^2 = 2x^2$, i.e., $x^4 - 4x^2 = 0$. Hence the points are $(0, 0), (2, 8), (-2, 8)$. It is understood that we have to find the area of the region given in Figure 1. The area is $\int_{-2}^{2}(f_2(x) - f_1(x))dx = \int_{-2}^{2}(2x^2 - x^4 + 2x^2)dx$.

2. Let us find the area bounded by the curves: $x = 3y - y^2$ and $x + y = 3$, i.e., $x = 3 - y$. The points of intersection are $(1, 1), (3, 0)$. Note that $(3y - y^2) - (3 - y) = -(y - 1)(y - 3) \geq 0$ for all $1 \leq y \leq 3$. Therefore the area is $\int_1^3 (3y - y^2) - (3 - y)dy$.



Polar Coordinates: To get a geometric idea we always relate a given function with a curve which is the graph of the given function. Sometimes we have to represent or express a given curve analytically (by a function or an equation). Expressing a given curve by the graph of a function or by an implicit equation using rectangular coordinates may not be always easy. Even if it is possible, in some cases, the function or the implicit equation may be complicated to use. Sometimes the polar coordinate system is better suited for the representation of a curve given geometrically. The term "curve" appearing here is the one which we usually imagine intuitively.

The polar coordinates are defined as follows. In the plane, we fix an origin O and an initial ray from O as shown in Figure 2. Then each point P in the plane can be assigned polar coordinates (r, θ) where r is the directed distance from O to P and θ is the directed angle from the initial ray to the segment OP.

The meaning of the directed angle is that the angle θ is positive when measured counterclockwise and negative when measured clockwise. The directed distance is something new. We will explain this concept with an example. Consider the points P and Q given in Figure 2. Here we assume that the lengths OP and OQ are same. Suppose $P = (5, 30)$, then Q is represented by $(-5, 30)$.

The negative distance can be understood as follows. If we go forward on the line QOP from O by the distance 5 we reach P and if we come backward on the same line from O by the distance 5 we reach Q. Note that Q has several representations, for example,

$$Q = (5, 210) = (-5, 30) = (5, -150) = (5, 570) = (-5, 390).$$

Of course one can ask what is the advantages of taking this directed distance (and the directed angle). We will take up the discussion on this question later.

Polar and Cartesian coordinates: If we use the common origin and take the initial ray as the positive x-axis, then the polar coordinates are related to the rectangular coordinates (x, y) by the equations:

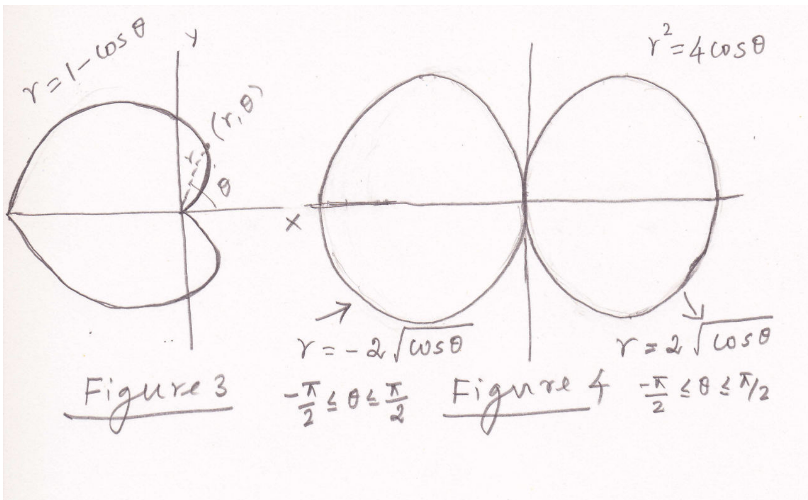
$$x = r\cos\theta,\ y = r\sin\theta \ \text{ or } \ x^2 + y^2 = r^2,\ \frac{y}{x} = \tan\theta.$$

Note that these equations are valid even if r is negative because $\cos(\theta+180) = -\cos\theta$, $\sin(\theta+180) = -\sin\theta$. The above equations are used to find Cartesian equations equivalent to polar equations and vice versa.

Example : $r^2 = 3r\sin\theta$ is equivalent to $x^2+y^2 = 3y$ which is a circle and $r\cos\theta = -4$ is equivalent to $x = -4$ which is a vertical line.

Graphs of the Polar Equations: A simple equation such as $r = 0$ (resp., $r = a$, $r = -a$, $\theta = \alpha$) represents the origin (resp., circle, the same circle, a straight line). We will now see how to represent the graph of a function given in polar equation: $r = f(\theta)$ or $F(r, \theta) = 0$. If the polar equation is given as $r = f(\theta)$, for sketching, we substitute a value of θ and find the corresponding $r = f(\theta)$. Then we plot the point (r, θ). To plot the curve we plot few points corresponding to few $\theta's$. To get the actual shape of the curve, it is desirable to consider the $\theta's$ for which $f(\theta)$ is a maximum or a minimum. As we do in the Cartesian case it is also desirable to consider the symmetry.

For example, the curve is symmetric about the origin (resp., x-axis, y-axis) if the equation is unchanged when r is replaced by $-r$ (resp., $-\theta$, $\pi-\theta$). The curve is also symmetric about the origin if the equation is unchanged when θ is replaced by $\theta + \pi$. Similarly, the curve is also symmetric about the x-axis (resp. y-axis) if the equation is unchanged when the pair (r, θ) is replaced by the pair $(-r, \pi - \theta)$ (resp., $(-r, -\theta)$).

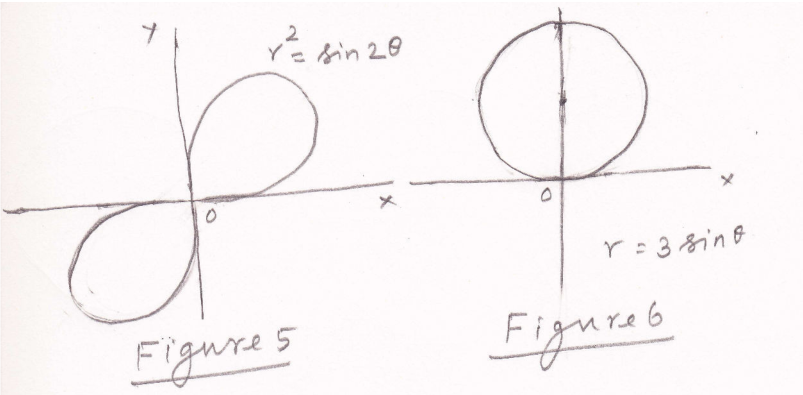


Examples: 1. Let us sketch the curve $r = f(\theta) = a(1 - \cos\theta), a > 0$. Since $\cos(-\theta) = \cos\theta$. the curve is symmetric about the x-axis. We note that $0 \le r = f(\theta) \le 2a$ and $r = 0$ occur at $\theta = 0$ and $r = 2a$ occurs at $\theta = \pi$. Moreover, $1 - \cos\theta$ increases from 0 to 2. For $\theta = \pi/3$ and $\pi/2$, we have $r = a$ and $a/2$ respectively. With this information, we can plot the curve (see Figure 3).

2. Consider the equation $r^2 = 4\cos\theta$. This equation is not given in the form $r = f(\theta)$. The graph of this equation can be plotted in the following ways. By varying θ from 0 to $\pi/2$ we get the corresponding values of r. Since the curve is symmetric over the x-axis and y-axis, we get the curve as given in Figure 4. The other way is to convert the equation in the form $r = \pm 2\sqrt{\cos\theta}$ and sketch the graphs of the equations $r = +2\sqrt{\cos\theta}$ and $r = -2\sqrt{\cos\theta}$. We get one portion of the curve given in Figure 4 by plotting $(2\sqrt{\cos\theta}, \theta)$ for $-\pi/2 \le \theta \le \pi/2$ and the other by plotting $(-2\sqrt{\cos\theta}, \theta)$ for $-\pi/2 \le \theta \le \pi/2$. What would be the graph of the function $r^2 = -4\cos\theta$?

3. Consider the equation $r^2 = \sin 2\theta$. As we did in the previous example we can sketch the graph of $r = \pm\sqrt{sin 2\theta}$. Interestingly, in this case the graphs of $r = +\sqrt{\sin 2\theta}$ and $r = -\sqrt{\sin 2\theta}$ coincide. The graph is given in Figure 5.

4. Consider the equation $r = 3\sin\theta$. If we plot (r, θ) for $0 \le \theta \le \pi$, we get the curve given in Figure 6 and if we plot (r, θ) for $\pi \le \theta \le 2\pi$ we get the same curve.



Remarks: 1. A point (r, θ) may not satisfy the equation $r = f(\theta)$ or $F(r, \theta) = 0$, however, it may still lie on the graph of the equation. For example $(2, \pi/2)$ does not satisfy the equation $r = 2\cos 2\theta$, however, $(2, \pi/2)$ lies on the curve, because $(-2, -\pi/2) = (2, \pi/2)$ and $(-2, -\pi/2)$ satisfies the equation. So the only sure way to identify all the points of intersection of two graphs is to sketch the graphs. Because solving of two equations may not lead to identifying all their points of intersection. We will see an example in the next lecture.

2. We will be dealing with the polar equations and their graphs only in the next one or two lectures. Later we will mainly use the polar coordinates to change the variables x and y to r and θ. In such cases we will assume $r > 0$ and $\theta \in [0, 2\pi)$, (at least we do not have to deal with the directed distance).

3. Allowing r to be negative has some advantages. For example, we could express the curve given in Figure 4 in a simple equation $r^2 = 4\cos\theta$. Several curves, especially those curves which are symmetric over the origin or the x-axis (see the lemniscate given in Figure 5), can be expressed in simpler forms if we allow the negative distance.

Note that the Cartesian equation $(x^2 + y^2)(x^2 + y^2)^2 = 16x^2$ is equivalent to the polar equation $r^2 = 4\cos\theta$. If we plot the points $(x, y)'s$ satisfying the Cartesian equation, we can see the symmetries over x-axis, y-axis and the origin. In fact the curves represented by the above Cartesian and the polar equations are same.

Lecture 20: Area in Polar coordinates; Volume of Solids

We will define the area of a plane region between two curves given by polar equations.

Suppose we are given a continuous function $r = f(\theta)$, defined in some interval $\alpha \le \theta \le \beta$. Let us also assume that $f(\theta) \ge 0$ and $\beta \le \alpha + 2\pi$. We want to define the area of the region bounded by the curve $r = f(\theta)$ and the lines $\theta = \alpha$ and $\theta = \beta$ (see Figure 1). Consider a partition $P : \alpha = \theta_0 < \theta_1 < ... < \theta_n = \beta$. Corresponding to P, we consider the circular sectors (as shown in Figure 1) with radii $f(\theta_i)'s$. Note that the area of the union of these circular sectors is approximately equal to the area of the desired region. The area of the sector OTS is

$$\left(\frac{\theta_i - \theta_{i-1}}{2\pi}\right)\pi f(\theta_i)^2 = \frac{1}{2}\triangle\theta_i f(\theta_i)^2.$$

So the sum of the areas of all the sectors is $\sum_{i=1}^{n} \frac{1}{2} f(\theta_i)^2 \triangle\theta_i$. Since f is continuous, this Riemann sum converges to $\frac{1}{2}\int_\alpha^\beta f(\theta)^2 d\theta$. In view of this we define the area bounded by the curve $r = f(\theta)$ and the lines $\theta = \alpha$ and $\theta = \beta$ to be

$$\frac{1}{2}\int_\alpha^\beta f(\theta)^2 d\theta.$$

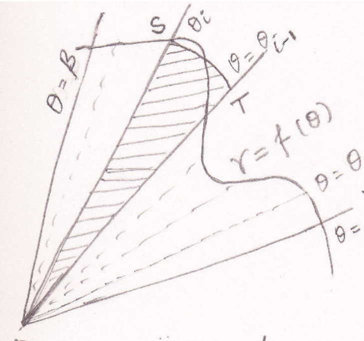


Figure 1

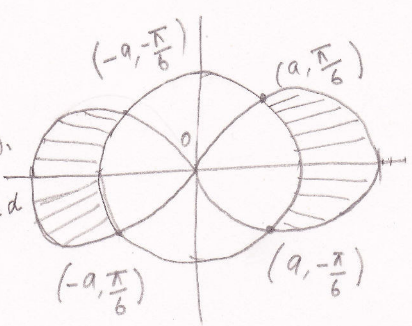


Figure 2

Example: Let us find the area inside the curve $r^2 = 2a^2\cos 2\theta$ and not included in the circle $r = a$. The graphs of the curves are given in Figure 2. To find the points of intersection we solve the equations which imply that $2a^2\cos 2\theta = a^2$. This implies that $\cos 2\theta = 1/2$ which in turn gives that $2\theta = \pi/3$. Therefore $\theta = \pi/6$.

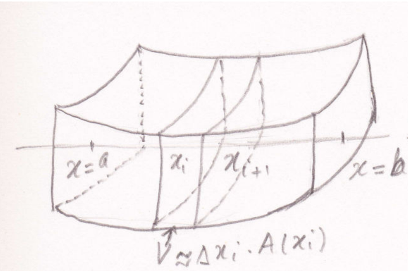


Figure 3

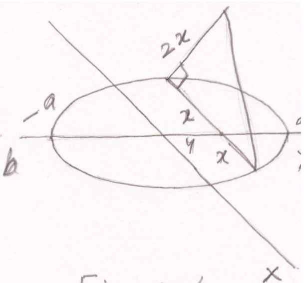


Figure 4

Using the symmetry we can get the other points of intersection. Actually we do not need the other points of intersection to solve this problem. The area of the desired region is

$$4\int_0^{\pi/6} \frac{1}{2} r^2 d\theta = 4\int_0^{\pi/6} \frac{1}{2}(2a^2\cos 2\theta - a^2) d\theta.$$

Volume of a solid by slicing: We will see that volumes of certain solid bodies can be defined as integral expressions.

Consider a solid which is bounded by two parallel planes perpendicular to x-axis at $x = a$ and $x = b$ (as shown in Figure 3). Let $P : x_0 = a < x_1 < x_2 < ... < x_n = b$ be a partition of $[a, b]$. For P consider the planes perpendicular to the x-axis at $x = x_i$'s. This would slice the given solid. Consider one slice of the solid that is bounded between two planes at $x = x_i$ and $x = x_{i+1}$ (see Figure 3). The volume of this slice is approximately $A(x_i)\triangle x_i$ where $A(x_i)$ is the area of the cross section of the solid made by the plane at $x = x_i$. Therefore it is natural to consider $\sum_{i=1}^{n} A(x_i)\triangle x_i$ as an approximation of the volume of the given solid. If $A(x)$ is continuous, then the above Riemann sum converges to $\int_a^b A(x)dx$. In view of the above, we define the volume of the solid to be

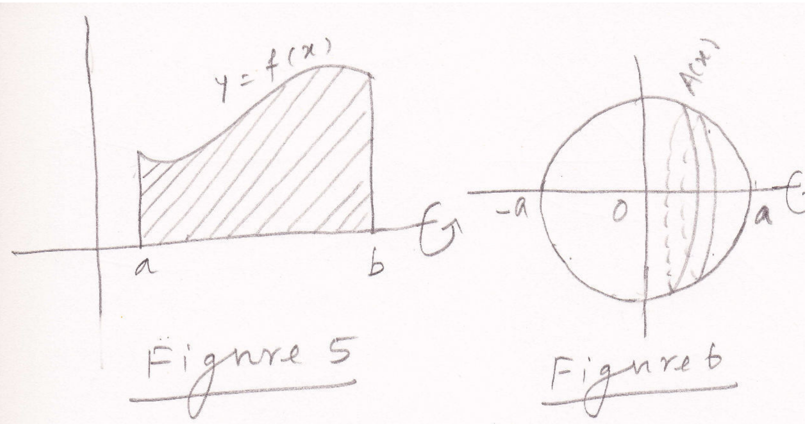
$$\int_a^b A(x)dx.$$

Example: The base of certain solid is the disk: $x^2 + y^2 \leq a^2$. Each section of the solid cut out by a plane perpendicular to the y-axis is an isosceles right triangle with one leg in the base of the solid (see Figure 4). Let us find the volume of the solid. By the above formula the volume is $V = \int_{-a}^{a} A(y)dy$ where $A(y)$ is the area of the cross section of the solid at y. Note that

$$A(y) = \frac{1}{2}(2x)^2 = 2(a^2 - y^2)dy.$$

Therefore, $V = \int_{-a}^{a} 2(a^2 - y^2)dy = \frac{8a^3}{3}$.

Volumes of solids of revolution: By revolving a planer region about an axis, we can generate a solid in $\mathbb{R}^3$. Such a solid is called a solid of revolution. We will use the method of finding volume by slicing and find the volume of a solid of revolution.



Consider a planer region which is bounded by the graph of a continuous function $f(x)$, $a \leq x \leq b$ where $f(x) \geq 0$ and the x-axis (see Figure 5). Suppose the region is revolved about the x-axis. The volume of the solid of revolution, by the slice method, is

$$V = \int_a^b A(x)dx = \int_a^b \pi\,(f(x))^2\,dx.$$

Example: For understanding, let us take a simple example of evaluating the volume of a sphere which is generated by the circular disk : $x^2 + y^2 \leq a^2$ by revolving it about the x-axis (see Figure 6). The volume is

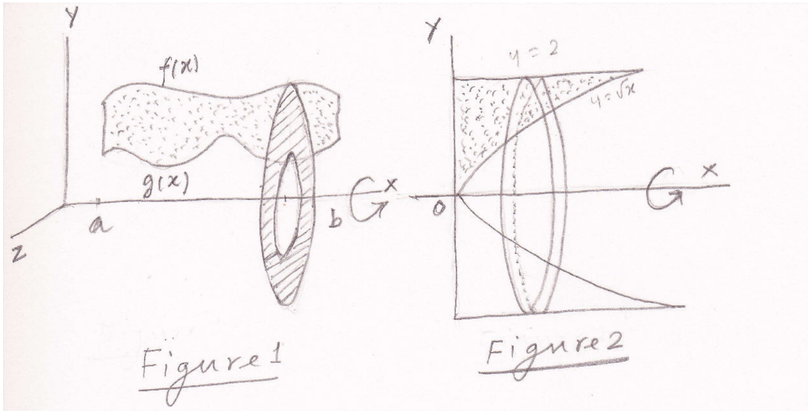
$$V = \int_{-a}^{a} \pi\,(f(x))^2\,dx = \int_{-a}^{a} \pi(a^2 - x^2)dx = \frac{4}{3}\pi a^3.$$

Lecture 21: Washer and Shell Methods; Length of a plane curve

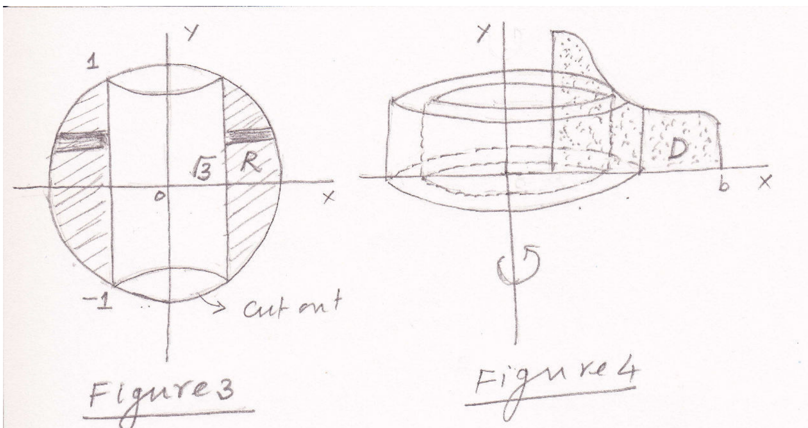
In the last lecture we considered the region between the graph of a continuous function $f(x),\ a \leq x \leq b$ where $f(x) \geq 0$ and the x-axis, and defined the volume of the solid generated by revolving this region about the x-axis. Instead of the x-axis we can take the graph of another function $g(x)$ such that $0 \leq g(x) \leq f(x),\ a \leq x \leq b$ and consider the region between the two graphs (see Figure 1). The volume of the solid generated by this region is

$$V = \int_a^b \pi(f(x)^2 - g(x)^2)dx.$$

In this case, the cross sections of the solid (i.e, the slice) perpendicular to the x-axis are washers instead of disks. Therefore this method of finding volume is called **washer method**.



Examples: 1. Let us find the volume generated by revolving the region bounded by $y = \sqrt{x},\ y = 2$ and $x = 0$ about the x-axis (see Figure 2). By washer method the volume is $V = \int_0^4 \pi(f(x)^2 - g(x)^2)dx$ where $f(x) = 2$ and $g(x) = \sqrt{x}$.



2. A round hole of radius $\sqrt{3}$ cms is bored through the center of a solid sphere of radius 2 cms. Let us find the volume cut out (see Figure 3). We know that the volume of the whole sphere is $\frac{4}{3}\pi r^3 = \frac{32}{3}\pi$. We will find the volume of the shaded region given in Figure 3 using the washer method. Note that the region R is revolved about the y-axis. We get the range for y by solving the equations $x^2+y^2 = 4$ and $x = \sqrt{3}$. By washer method the volume is $\int_{-1}^{1} \pi(f(y)^2 - 3)dy = \int_{-1}^{1} \pi(4 - y^2 - 3)dy = \frac{4\pi}{3}$. The volume cut out is $\frac{32\pi}{3} - \frac{4\pi}{3} = \frac{28\pi}{3}$.

Shell Method: Recall that in the washer method we consider the slices perpendicular to the axis of revolution which look like washers. We now describe another method, called shell method, in which we will consider the slices parallel to the axis of revolution which will look like shells.

Let D be a plane region between the graph of the function $f : [a, b] \to \mathbb{R},\ a > 0$, and the x-axis as shown in Figure 4. We define the volume of the solid generated by revolving D about the y-axis to be

$$V = \int_a^b 2\pi x f(x) dx.$$

This formula is motivated by the following fact. For $x \in [a, b]$, consider the small interval $I = [x - \triangle x/2, x + \triangle x/2]$ and the rectangle with the base I and the hight $f(x)$ (see Figure 4). If this rectangle is revolved about the y-axis, a cylindrical shell is generated and the shell's volume is $2\pi x f(x)\triangle x$.

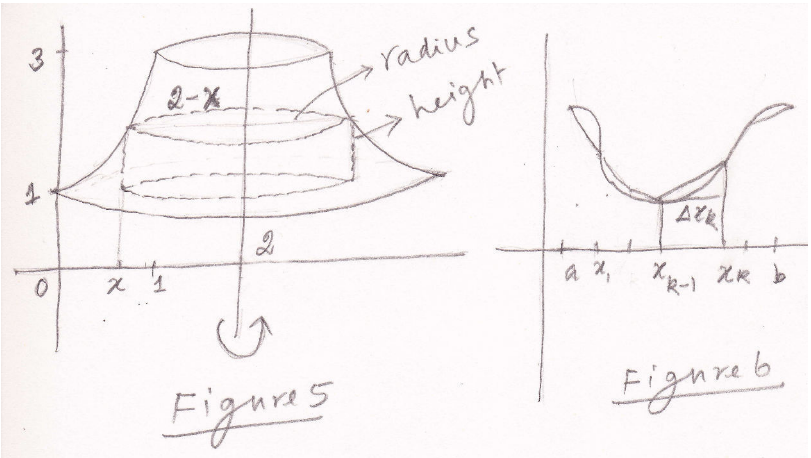
Example : Consider the solid obtained by revolving the region bounded by the functions

$$y = x^2 + x + 1,\ y = 1,\ \text{and}\ \ x = 1$$

about the line $x = 2$ (see Figure 5). Let us find the volume of the solid by the shell method. By the shell method, the volume is

$$V = \int_a^b 2\pi\ \text{(shell radius) (shell height)}\ dx.$$

For each x from 0 to 1, we consider a shell (see Figure 5). The shell radius at x is $2 - x$ and the shell height is $x^2 + x + 1 - 1$. Therefore the volume is $\int_0^1 2\pi\ (2 - x)\ (x^2 + x)dx$.



Remark: In the last two lectures we defined area of some region and volume of some solid in terms of some integral expressions. Note that we have not derived these integral expressions or the formulae. We have given some justifications for the definitions. For example, in the shell method, we approximated a shell by a cylindrical shell and justified the definition or the formula. Instead of the cylindrical approximation, if we take some other approximation we may end up with some other formula which may not give the same result. So in order to find the area of some region or volume of some solid one has to use the definition (or say the formula). One should not try to approximate the given region by some other region and derive the area or the volume in terms of some integral expression.

Length of a plane curve: Let $f : [a, b] \to \mathbb{R}$ be a differentiable function such that f' is continuous. Such a function is said to be smooth and its graph is said to be a smooth curve. We will describe the length of such a curve in terms of an integral expression.

Let $P : a = x_0 < x_1 < x_2 \ldots < x_n = b$ be a partition of $[a, b]$. Join $(x_{k-1}, f(x_{k-1}))$ and $(x_k, f(x_k))$ by a straight line (see Figure 6). Then we define the length L of the curve to be

$$L = \lim_{\|P\| \to 0} \sum_{k=1}^{n} \sqrt{(\triangle x_k)^2 + (f(x_k) - f(x_{k-1}))^2}.$$

By MVT, there exists $c_k \in (x_{k-1}, x_k)$ such that $f(x_k) - f(x_{k-1}) = f'(c_k)\triangle x_k$. Therefore

$$L = \lim_{\|P\|\to 0} \sum_{k=1}^{n} \sqrt{(\triangle x_k)^2 + f'(c_k)(\triangle x_k)^2} = \lim_{\|P\|\to 0} \sum_{k=1}^{n} \sqrt{1 + f'(c_k)^2}\triangle x_k = \int_a^b \sqrt{1 + f'(x)^2} dx.$$

Length of a plane curve given in parametric form:

Suppose a plane curve is given by $\{((x(t), y(t)) : t \in [a, b]\}$ where x and y are continuous function. Such a curve is called a parametric curve. For example unit circle is given by $\{(\cos t, \sin t) : t \in [0, 2\pi]\}$. We assume that x' and y' are continuous. Such a curve is said to be smooth. For a smooth curve the length of the curve L is defined to be

$$L = \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2}\ dt.$$

Example: Consider the curve is given by $\{(a\cos^3 t, a\sin^3 t) : t \in [0, 2\pi]\}$ for some $a > 0$. The length of this curve is

$$L = \int_0^{2\pi} \sqrt{(-3a\cos^2 t\sin t)^2 + (3a\sin^2 t\cos t)^2}\ dt = \int_0^{2\pi} 3a \mid \cos t \sin t \mid dt = 4\frac{3a}{2}\int_0^{\frac{\pi}{2}} \sin 2t dt.$$

Curve given in polar form:

If a curve is given in the polar form $r = f(\theta), f(\theta) \geq 0,\ \alpha \leq \theta \leq \beta$ then the curve can be represented in parametric form by taking $x(\theta) = r(\theta)\sin\theta$ and $y(\theta) = r(\theta)\cos\theta$. In this case

$$L = \int_\alpha^\beta \sqrt{(d(r\cos\theta))^2 + (d(r\sin\theta))^2}\ d\theta = \int_\alpha^\beta \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\ d\theta.$$

Example: Consider the cardioid $r = a(1 + \cos\theta)$, where $a > 0,\ 0 \leq \theta \leq 2\pi$. The length of the curve is

$$L = \int_0^{2\pi} \sqrt{2a^2(1 + \cos\theta)}\ d\theta = 2a\int_0^{2\pi} \mid \cos(\theta/2) \mid d\theta = 4a\int_0^{\pi} \cos(\theta/2) d\theta = 8a.$$

Problem 1: *The region bounded by the curve $y = x^2 + 1$ and the line $y = -x + 3$ is revolved about the X-axis to generate a solid. Find the volume of the solid.*

Solution: We use the washer method. The outer radius $r_2(x) = -x + 3$ and the inner radius $r_1(x) = x^2 + 1$. Therefore, the volume is

$$V = \int_{-2}^{1} \pi((-x + 3)^2 - (x^2 + 1)^2)dx = 117\frac{\pi}{3}.$$

Problem 2: *The region in the first quadrant bounded by the parabola $y = x^2$, the Y-axis and the line $y = 1$ is revolved about the line $x = 2$ to generate a solid. Find the volume of the solid.*

Solution: We use the shell method. Observe that shell radius is $2 - x$ and the shell height is $1 - x^2$. The volume $V = \int_0^1 2\pi(2 - x)(1 - x^2)dx = \frac{13\pi}{6}$.

The volume can also be computed by the washer method.

Lecture 22: Areas of surfaces of revolution, Pappus's Theorems

Let $f : [a, b] \to \mathbb{R}$ be continuous and $f(x) \geq 0$. Consider the curve C given by the graph of the function f. Let S be the surface generated by revolving this curve about the x-axis. We will define the surface area of S in terms of an integral expression.

Consider a partition $P : a = x_0 < x_1 < x_2 < ... < x_n = b$ and consider the points $P_i = (x_i, f(x_i)), i = 0, 1, 2, .., n$. Join these points by straight lines as shown in Figure 1. Consider the segment $P_{i-1}P_i$. The area A of the surface generated by revolving this segment about the x-axis is $\pi(f(x_{i-1}) + f(x_i))\ell_i$ where ℓ_i is the length of the segment $P_{i-1}P_i$. This can be verified as follows. Note that the area $A = \pi f(x_i)(\ell + \ell_i) - \pi f(x_{i-1})\ell$ (see Figure 2). Since

$$\frac{\ell}{f(x_{i-1})} = \frac{\ell + \ell_i}{f(x_i)} = \frac{\ell_i}{f(x_i) - f(x_{i-1})} = \alpha$$

for some α, the area

$$A = \pi f(x_i)\alpha f(x_i) - \pi f(x_{i-1})\alpha f(x_{i-1}) = \pi\alpha(f(x_i) + f(x_{i-1}))(f(x_i) - f(x_{i-1})) = \pi\ell_i(f(x_{i-1}) + f(x_i)).$$

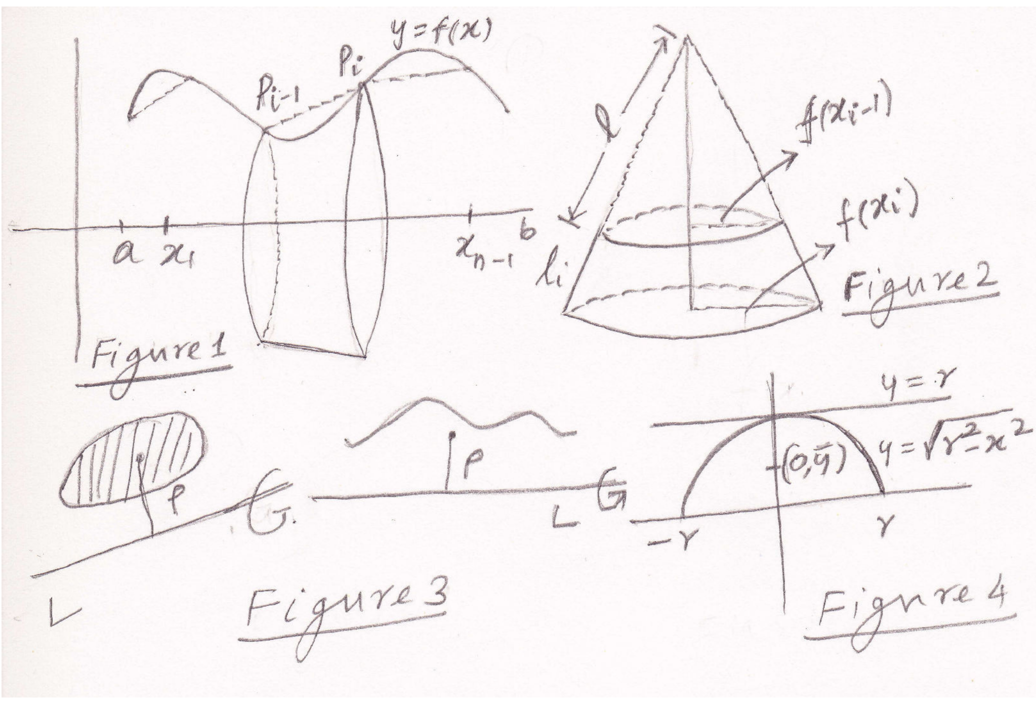
The sum of the areas of the surfaces generated by the line segments is

$$\sum_{i=1}^{n} \pi(f(x_{i-1}) + f(x_i))\ell_i = \sum_{i=1}^{n} \pi f(x_{i-1})\sqrt{(\triangle x_i)^2 + (\triangle y_i)}^2 + \sum_{i=1}^{n} \pi f(x_i)\sqrt{(\triangle x_i)^2 + (\triangle y_i)}^2$$

where $\triangle y_i = f(x_i) - f(x_{i-1})$. If f' is continuous, one can show that each of the sum given in the RHS of the above equation converges to $\int_a^b \pi f(x)\sqrt{1 + (f'(x))^2}\, dx$ as $\| P \| \to 0$. In view of this we define the surface area generated by revolving the curve about the x-axis to be

$$\int_a^b 2\pi f(x)\sqrt{1 + (f'(x))^2}\, dx.$$

In case $f(x) \leq 0$, the formula for the area is $\int_a^b 2\pi \mid f(x) \mid \sqrt{1 + (f'(x))^2}\, dx$.



Example: Let us find the area of the surface generated by revolving the curve $y = \frac{1}{2}(x^2 + 1), 0 \leq x \leq 1$ about the y-axis. Here the function y is increasing hence it is one-one and onto. Hence we can write x in terms of y : $x = g(x) = \sqrt{2y-1}$. In this case the formula is $\int_a^b 2\pi \mid g(y) \mid \sqrt{1+(g'(y))^2}\, dy$ where $a = 1/2$ and $b = 1$.

Parametric case: If the curve is given in the parametric form $\{(x(t), y(t)) : t \in [a,b]\}$, and x' and y' are continuous, then the surface area generated is

$$\int_a^b 2\pi\rho(t)\sqrt{(x'(t))^2+(y'(t))^2}\, dt$$

where $\rho(t)$ is the distance between the axis of revolution and the curve.

Example : The curve $x = t+1,\ y = \frac{t^2}{2}+t,\ 0 \le t \le 4$ is rotated about the y-axis. Let us find the surface area generated. The surface area is $\int_0^4 2\pi \mid t+1 \mid \sqrt{1+(1+t)^2}dt$.

Polar case: If the curve is given in the polar form, the surface area generated by revolving the curve about the x-axis is

$$\int_a^b 2\pi y\sqrt{r^2+\left(\frac{dr}{d\theta}\right)^2}\, d\theta = \int_a^b 2\pi\ r(\theta)\sin\theta\ \sqrt{r^2+\left(\frac{dr}{d\theta}\right)^2}\, d\theta.$$

Example : The lemniscate $r^2 = 2a^2\cos 2\theta$ is rotated about the x-axis. Let us find the area of the surface generated. A simple calculation shows that $\sqrt{r^2+\left(\frac{dr}{d\theta}\right)^2} = \frac{2a^2}{r}$. The curve is given in the notes of the previous lecture. The surface area is $2\int_0^{\frac{\pi}{4}} 2\pi r\sin\theta\frac{2a^2}{r}d\theta = 8\pi a^2(1-\frac{1}{\sqrt{2}})$.

Pappus's Theorems: There are two results of Pappus which relate the centroids to surfaces and solids of revolutions. The first result relates the centroid of a plane region with the volume of the solid of revolution generated by it.

Theorem: *Let R be a plane region. Suppose R is revolved about the line L which does not cut through the interior of R, then the volume of the solid generated is*

$$V = 2\pi\rho A$$

where ρ is the distance from the axis of revolution to the centroid and A is the area of the region R(see Figure 3).

Note that in the above formula $2\pi\rho$ is the distance traveled by the centroid during the revolution. The second result relates the centroid of a plane curve with the area of the surface of revolution generated by the curve.

Theorem: *Let C be a plane curve. Suppose C is revolved about the line L which does not cut through the interior of C, then the area of the surface generated is*

$$S = 2\pi\rho L$$

where ρ is the distance from the axis of revolution to the centroid and L is the length of the curve C (see Figure 3).

Example: Use a theorem of Pappus to find the centroid of the semi circular arc $y = \sqrt{r^2-x^2},\ -r \le x \le r$. If the arc is revolved about the line $y = r$, find the volume of the surface area generate.

Solution: We know the surface area generated by the curve $4\pi r^2$ (see Figure 4). Let the centroid of the curve be $(0, \overline{y})$. By Pappus theorem $4\pi r^2 = 2\pi\overline{y}\pi r$ which implies that $\overline{y} = \frac{2r}{\pi}$. Again by Pappus theorem, the area of the surface generated by revolving the curve around $y = r$ is $2\pi(r-\overline{y})\pi r = 2\pi r^2(\pi-2)$.

Lecture 23: Review of vectors, equations of lines and planes; Sequences in $\mathbb{R}^3$

In the next two lectures we will deal with the functions from $\mathbb{R}$ to $\mathbb{R}^3$. Such functions are called vector valued functions. After two lectures we will deal with the functions of several variables, that is, functions from $\mathbb{R}^3$ or $\mathbb{R}^n$ to $\mathbb{R}$. Before discussing about the functions let us see some properties of $\mathbb{R}^3$. We first review some basic concepts from vector algebra.

Norm of a vector: If $X = (x, y, z)$, then the norm of X, denoted by $\|X\|$, is $\sqrt{x^2 + y^2 + z^2}$. $\|X - Y\|$ is the distance between the points X and Y.

Scalar product of two vectors: If $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$, then the scalar product of X and Y is $X \cdot Y = x_1y_1 + x_2y_2 + x_3y_3$.

Projection of a vector: The projection of a vector A along the non-zero vector B is $\frac{A \cdot B}{B \cdot B}B$.

Angle between two vectors : If θ is the angle between two vectors A and B then $A \cdot B = \|A\|\|B\| \cos\theta$.

Parametric and Cartesian equations of straight lines: The parametric representation of the straight line passing through P and parallel to a (non-zero) vector is $X - P = tA,\ t \in \mathbb{R}.$. If $X = (x, y, z),\ P = x_0, y_0, z_0)$ and $A = (a, b, c)$, then the above equation becomes

$$x = x_0 + ta,\ y = y_0 + tb \text{ and } z = z_0 + tc.$$

In case $a, b, c \neq 0$, then the equation of the line is

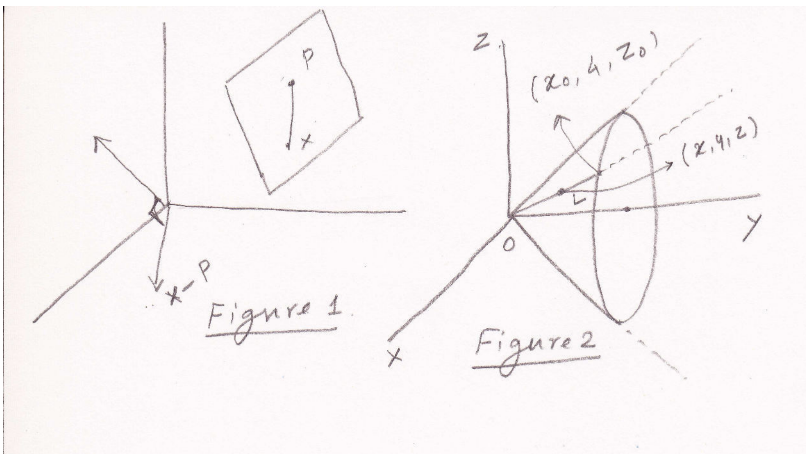
$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

If $a = 0$, then the line is represented as $x = x_0$ and $\frac{y-y_0}{b} = \frac{z-z_0}{c}$.

Equation of a plane (passing through a point and perpendicular to a vector): The set of points $\{X : (X - P) \cdot N\} = \{X : X \cdot N = P \cdot N\}$ in $\mathbb{R}^3$ is the plane perpendicular to the vector N and passing the point P (see Figure 1). If $N = (a, b, c)$ and $P = (x_0, y_0, z_0)$ then the equation of the plane is

$$(x, y, z) \cdot (a, b, c) = (x_0, y_0, z_0) \cdot (a, b, c),$$

that is, $ax + by + cz = ax_0 + by_0 + cz_0$.



We can also derive equations of certain surfaces similar to the way we derived the equation of a plane.

Example: *Find the equation of the right circular cone having vertex at the origin and passing through the circle* $x^2 + y^2 = 25, \quad y = 4.$.

Solution: Let (x, y, z) be any arbitrary point on the surface. Let L be the straight line passing through (x, y, z) and $(0, 0, 0)$. Let $(x_0, 4, z_0)$ be the point of intersection of the line and the circle (see Figure 2). The equation of the line L is $\frac{x}{x_0} = \frac{y}{4} = \frac{z}{z_0}$. This implies that $x_0 = 4x/y$ and $z_0 = 4z/y$. Since x_0 and z_0 satisfy the equation of the circle, we have $4^2(\frac{x}{y})^2 + 4^2(\frac{z}{y})^2 = 25$. This implies that $16(x^2 + z^2) = 25y^2$.

Problem : *Determine the equation of the cylinder generated by a line through the curve* $(x-2)^2 + y^2 = 4,\ z = 0$ *moving parallel to the vector* $\vec{i} + \vec{j} + \vec{k}$.

Solution: Any point on the curve is of the form $(x_0, y_0, 0)$. The equation of a line passing through $(x_0, y_0, 0)$ and parallel to $(1, 1, 1)$ is $\frac{x-x_0}{1} = \frac{y-y_0}{1} = \frac{z}{1}$. We get $x_0 = x - z$ and $y_0 = y - z$. Since $(x_0, y_0, 0)$ lies on the curve, we get the equation of the cylinder to be $(x - z - 2)^2 + (y - z)^2 = 4$.

Convergence of a sequence in $\mathbb{R}^3$: We will see that the concept of convergence of sequence in $\mathbb{R}^3$ plays a role in studying about the vector valued functions and functions of several variables.

Let $X_n = (x_{1,n}, x_{2,n}, x_{3,n}) \in \mathbb{R}^3$. We say that the sequence (X_n) is convergent if there exists $X_0 \in \mathbb{R}^3$ such that $\|X_n - X_0\| \to 0$ as $n \to \infty$. In this case we say that X_n converges to X_0 and we write $X_n \to X_0$.

Note that corresponding to a sequence (X_n), $X_n = (x_{1,n}, x_{2,n}, x_{3,n})$, there are three sequences $(x_{1,n})(x_{2,n})$ and $(x_{3,n})$ in $\mathbb{R}$, and vice-versa. We will see that the properties of (X_n) can be completely understood in terms of the properties of the corresponding sequences $(x_{1,n})(x_{2,n})$ and $(x_{3,n})$ in $\mathbb{R}$.

Theorem 1. $X_n \to X_0$ *in* $\mathbb{R}^3 \Leftrightarrow$ *the coordinates* $x_{i,n} \to x_{i,0}$ *for every* $i = 1, 2, 3$ *in* $\mathbb{R}$.

Proof: This follows from the fact that $\sum_{i=1}^{3} |x_{i,n} - x_{i,0}|^2 \to 0 \Leftrightarrow |x_{i,n} - x_{i,0}| \to 0,\ i = 1, 2, 3.$ $\square$

The proof of the following result is similar to the proof of the previous result.

Theorem 2. (X_n) is bounded (i.e., $\exists\ M$ such that $\|X_n\| \leq M\ \forall\ n$) $\Leftrightarrow$ each sequence $(x_{i,n}), i = 1, 2, 3$, is bounded.

Problem 1: *Every convergent sequence in* $\mathbb{R}^3$ *is bounded.*

Proof: If $\|X_n - X_0\| \to 0$, then $(\|X_n - X_0\|)$ is bounded. This implies that $(\|X_n\|)$ is bounded and this proves the result. $\square$

Problem 2 (Bolzano-Weierstrass Theorem): *Every bounded sequence in* $\mathbb{R}^2$ *has a convergent subsequence.*

Proof (*): Suppose (x_n, y_n) be a bounded sequence. By Theorem 2 both (x_n) and (y_n) are bounded. By B-W theorem (x_n) has a convergent subsequence, say $x_{n_k} \to x_0$. Consider the sequence (y_{n_k}) and note that this sequence is also bounded. Again by B-W theorem, this sequence has a convergent subsequence, say $y_{n_{k_i}} \to y_0$. It is clear that the subsequence $(y_{n_{k_i}}, y_{n_{k_i}})$ of (x_n, y_n) converges to (x_0, y_0). $\square$.

It is evident that the above theorem can also be extended to $\mathbb{R}^3$.

Lecture 24 : Calculus of vector valued functions

In the previous lectures we had been dealing with functions from a subset of $\mathbb{R}$ to $\mathbb{R}$. In this lecture we will deal with the functions whose domain is a subset of $\mathbb{R}$ and whose range is in $\mathbb{R}^3$ (or $\mathbb{R}^n$). Such functions are called vector valued functions of a real variable.

If the values of a function F are in $\mathbb{R}^3$, then each $F(t)$ has 3 components, for example $F(t) = ((f_1(t), f_2(t), f_3(t))$. Therefore, each vector valued function F is associated with 3 real valued functions f_1, f_2 and f_3 and in this case we write $F = (f_1, f_2, f_3)$.

Let us see some examples of vector valued functions.

Examples: 1. Let $X_0, P \in \mathbb{R}^3$ and $P \neq 0$. Consider the vector valued function $F(t) = X_0 + tP$. It is clear that the range of the vector valued function is the line though the point X_0 parallel to the vector P.

2. Consider the vector valued functions $F_1(t) = (cost, sint),\ 0 \leq t \leq 2\pi$ and $F_2(t) = (cost, sint, t)$, $\infty < t < \infty$. We can geometrically visualize the ranges of F_1 and F_2 as t varies. In fact $F_1(t)$ varies on a circle and $F_2(t)$ varies on a helix. Both these curves are particular cases of parametric curves.

Parametric curves: Let I be an interval and $F : I \to \mathbb{R}^3$. The set of points $\{F(t) : t \in I\}$ is called the graph of the function F. If F is continuous (for the definition see below) then such a graph is called a curve or parametric curve with the parameter t.

From the previous definition it is clear that each continuous vector valued function corresponds to a curve. Naturally one expects that some geometric properties of the curves can be investigated by using some properties of the vector valued functions.

In this lecture we will extend the basic concepts of calculus, such as limit, continuity and derivative, to vector valued functions and see some applications to the study of curves.

Limits and derivatives:

Let $F = (f_1, f_2, f_3)$ be a vector valued function and $L = (l_1, l_2, l_3)$.

We say that $\lim_{t\to t_0} F(t) = L$ if $\lim_{t\to t_0} \| F(t) - L \| = 0$.

Proposition: $\lim_{t\to t_0} F(t) = L$ if and only if $\lim_{t\to t_0} f_i(t) = l_i$ for $i = 1, 2, 3$.

Proof: This follows from the fact that $\sum_{i=1}^{3} \mid f_i(t) - l_i \mid^2 \to 0 \ \Leftrightarrow \ \mid f_i(t) - l_i \mid \to 0,\ i = 1, 2, 3.$ $\square$

From the previous result it follows that $\lim_{t\to t_0} F(t) = (\lim_{t\to t_0} f_1(t), \lim_{t\to t_0} f_2(t), \lim_{t\to t_0} f_3(t))$ whenever the component on the right is meaningful.

We say that F is continuous at t_0 if $\lim_{t\to t_0} F(t) = F(t_0)$. One can show that F is continuous at t_0 if and only if each of the component function f_i is continuous at t_0.

We say that F is differentiable at t_0 if $\lim_{h\to 0} \frac{F(t_0+h)-F(t_0)}{h}$ exists. The limit is called the derivative of F at t_0 and is denoted by $F'(t_0)$. Note that F is differentiable at t_0 if and only if f_i is differentiable at t_0 for all $i = 1, 2, 3$. Moreover, $F'(t_0) = (f_1'(t_0), f_2'(t_0), f_3'(t_0))$.

Tangent Vector: As in the case of a real valued function, we will see that the derivative $F'(t_0)$ is related to the concept of tangency. Suppose F is differentiable at t_0 and $F'(t_0) \neq 0$. Then

$$F'(t_0) = \lim_{h\to 0} \frac{1}{h}(F(t_0 + h) - F(t_0)).$$

Geometrically, one can visualize that the vector $\frac{1}{h}(F(t_0+h)-F(t_0))$, which is parallel to the vector $F(t_0+h)-F(t_0)$, moves to be a tangent vector as $h \to 0$. In view of this we have the following definition.

Definition: Suppose C is a curve defined by a differentiable vector valued function R. Suppose $R'(t_0) \neq 0$. The vector $R'(t_0)$ is called a tangent vector to C at $F(t_0)$ and the line $X(t) = R(t_0) + tR'(t_0)$ is called the tangent line to C at $R(t_0)$.

Example: Let us find the equation of the plane perpendicular to the circular helix $R(t) = (\cos t, \sin t, t)$ at $t_0 = \frac{\pi}{3}$. The equation of the plane passing through $R(\frac{\pi}{3})$ and perpendicular to $R'(\frac{\pi}{3})$ is the required plane. So the plane is $R'(\frac{\pi}{3}) \cdot (x, y, z) = R'(\frac{\pi}{3}) \cdot R(\frac{\pi}{3})$.

Arc length for space curves: We have seen a formula for evaluating the length of a plane curve. The formula can be extended to the space curves. Let C be a space curve defined by $R(t) = x(t)i + y(t)j + z(t)k$, $a \leq t \leq b$. Throughout this lecture we will assume that R' is continuous. The length of the curve C is defined to be

$$L = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}dt = \int_a^b \parallel \frac{dR}{dt} \parallel .$$

Arc length parameter: Let $R(t_0)$ be a fixed point on the curve C. For t, the directed distance measured along C from $R(t_0)$ and up to $R(t)$ is $s(t) = \int_{t_0}^t \sqrt{x'(\tau)^2 + y'(\tau)^2 + z'(\tau)^2}d\tau$. Each value of s corresponds to a point on C and this parametrizes C with respect to s, the arc length parameter. By the first FTC we have,

$$\frac{ds}{dt} = \parallel \frac{dR}{dt} \parallel .$$

This is expected. Because, if we consider $R(t)$ is the position vector of a particle moving along C, then $v(t) = R'(t)$ is the velocity vector and $a(t) = v'(t)$ is the acceleration vector. The speed with which the particle moves along its path is the magnitude of v.

Unit tangent vector: The unit tangent vector of $R(t)$ is $T = \frac{R'(t)}{\|R'(t)\|}$ whenever $\parallel R'(t) \parallel \neq 0$. From the derivation of $\frac{ds}{dt}$, we get $T = \frac{\frac{dR}{dt}}{\frac{ds}{dt}}$. Now can we write

$$T = \frac{dR}{dt}\frac{dt}{ds} = \frac{dR}{ds} \quad ?$$

The second equality of the above equation follows from the chain rule and the first equation follows from the following theorem.

Theorem: Let $f : [a, b] \to \mathbb{R}$, $f'(x) \neq 0$ for all $x \in [a, b]$. Then f^{-1} is continuous, differentiable and $\left(f^{-1}\right)'(f(x)) = \frac{1}{f'(x)}$.

Proof (*): Use the sequential argument and Bolzano-Weierstrass theorem to prove that f^{-1} is continuous. Let $f : [a, b] \to [c, d]$, $c \leq y_0 \leq d, y_0 = f(x_0)$ and $y = f(x)$ for some $x \in [a, b]$. Suppose $y \neq y_0$. Then,

$$\frac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} = \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{\frac{f(x)-f(x_0)}{x-x_0}}$$

Now let $y \to y_0$ and use the continuity of f^{-1} to get the result. □

Let us go back to the question we asked above. Let us work with the curve C such that $s(t)$ increases as $t > t_0$ increases, that is $\frac{ds}{dt} > 0$. By the previous theorem we have, $\frac{dt}{ds} = \frac{1}{\frac{ds}{dt}}$.

Lecture 25 : Principal Normal and Curvature

In the previous lecture we defined unit tangent vectors to space curves. In this lecture we will define normal vectors.

Consider the following results.

Theorem: If F and G are differentiable vector valued functions then so is $F \cdot G$ and $(F \cdot G)' = F' \cdot G + F \cdot G'$.

Proof : Let $F = (f_1, f_2, f_3)$ and $G = (g_1, g_2, g_3)$. Then $(F \cdot G)' = (f_1 g_1 + f_2 g_2 + f_3 g_3)'$. By simplifying this we get the result. $\square$

Theorem: Let I be an interval and F be a vector valued function on I such that $\| F(t) \| = \alpha$ for all $t \in I$. Then $F \cdot F' = 0$ on I, that is $F'(t)$ is perpendicular to $F(t)$ for each $t \in I$.

Proof: Let $g(t) = \| F(t) \|^2 = F(t) \cdot F(t)$. By assumption g is constant on I and therefore $g' = 0$ on I. By the previous theorem, $g' = 2F \cdot F'$. Therefore, $F \cdot F' = 0$. $\square$

Since the unit tangent vector T has constant length 1, by the previous theorem, T' is perpendicular to T. In view of this, we define the principle normal to the curve

$$N(t) = \frac{T'(t)}{\| T'(t) \|}$$

whenever, $\| T'(t) \| \neq 0$.

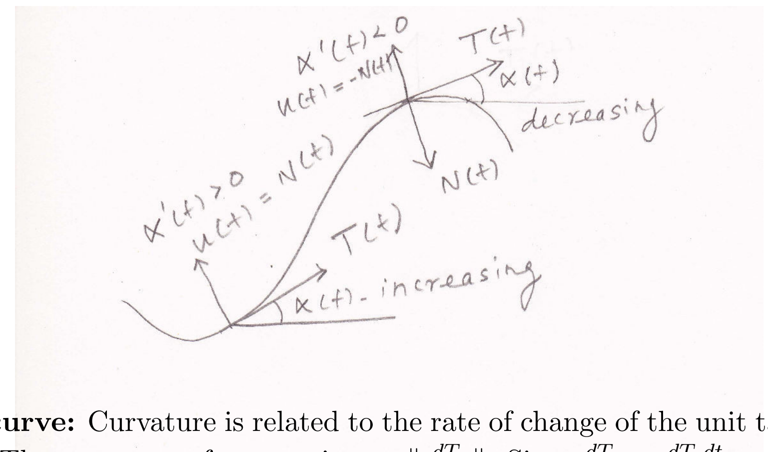
Geometric Interpretation: Let us consider a plane curve. Since T is a unit vector,

$$T(t) = \cos \alpha(t) i + \sin \alpha(t) j$$

where $\alpha(t)$ is the angle between the tangent vector and the positive x-axis (see the figure). From the previous equation, we get

$$T'(t) = -\sin \alpha(t) \alpha'(t) i + \cos \alpha(t) \alpha'(t) j = \alpha'(t) u(t)$$

where $u(t) = \cos(\alpha(t) + \frac{\pi}{2}) i + \sin(\alpha(t) + \frac{\pi}{2}) j$ which is a unit vector. When $\alpha'(t) > 0$, the angle is increasing and in this case $N(t) = u(t)$. Similarly, when $\alpha'(t) < 0$, the angle is decreasing and we have $N(t) = -u(t)$.



Curvature of a curve: Curvature is related to the rate of change of the unit tangent with respect to the arc length. The curvature of a curve is $\kappa = \| \frac{dT}{ds} \|$. Since $\frac{dT}{ds} = \frac{dT}{dt}\frac{dt}{ds} = \frac{T'(t)}{\| \frac{dR}{dt} \|}$, the curvature is given by the following formula:

$$\kappa(t) = \frac{\| T'(t) \|}{\| \frac{dR}{dt} \|}.$$

Example 1: Suppose C is a circle of radius a defined by $R(t) = a\cos ti + a\sin tj$. This implies that $R'(t) = -a\sin ti + a\cos tj$, $T(t) = -\sin ti + \cos tj$ and $T'(t) = -\cos ti - \sin tj$. Since $\| R'(t) \| = a$ and $\| T'(t) \| = 1$, we have $\kappa = \frac{1}{a}$. So the circle has the constant curvature and the curvature is the reciprocal of the radius of the circle.

Example 2: Sometimes the curvature of a plane curve is defined to be the rate of change of the angle between the tangent vector and the positive x-axis. We will see that our definition coincides with this. For a plane curve, we have shown that $\| T'(t) \| = | \alpha'(t) |$, when $T(t) = \cos\alpha(t)i + \sin\alpha(t)j$. By the chain rule, $\frac{d\alpha}{dt} = \frac{d\alpha}{ds}\frac{ds}{dt} = \| \frac{dR}{dt} \| \frac{d\alpha}{ds}$. From the formula of curvature, we get

$$\kappa(t) = | \frac{d\alpha}{ds} | .$$

The formula given in the following theorem provides a simpler method for determining the curvature.

Theorem: Let $v(t)$ and $a(t)$ denote the velocity and the acceleration vectors of a motion of a particle on a curve defined by $R(t)$. Then

$$\kappa(t) = \frac{\| a(t) \times v(t) \|}{\| v(t) \|^3}.$$

We will not prove the previous theorem but we will use it.

Problem 1: If a plane curve has the Cartesian equation $y = f(x)$ where f is a twice differentiable function, then show that the curvature at the point $(x, f(x))$ is $\dfrac{|f''(x)|}{[1 + f'(x)^2]^{3/2}}$.

Solution: The graph of f can be considered as a parametric curve $R(t) = ti + f(t)j$. Then $v(t) = R'(t) = i + f'(x)j$ and $a(t) = R''(t) = f''(t)j$. This implies that $v(t) \times a(t) = f''(t)k$. Therefore, $\| v(t) \times a(t) \| = | f''(t) |$ and $\| v(t) \| = \sqrt{1 + f'(t)^2}$. Substituting these values in the formula of $\kappa(t)$, we get the final expression.

Problem 2: For the curve $R(t) = t\vec{i} + t^2\vec{j} + \frac{2}{3}t^3\vec{k}$ find the equations of the tangent, principal normal and binormal at $t = 1$. Also calculate the curvature of the curve.

Solution : Differentiating $R(t)$ we get, $R'(t) = i + 2tj + 2t^2k$. The unit tangent vector is given by

$$T(t) = \frac{R'(t)}{\| R'(t) \|} = \frac{i + 2tj + 2t^2k}{1 + 2t^2}.$$

Differentiating, we get

$$T'(t) = \frac{-4ti + (2 - 4t^2)j + 4tk}{(1 + 2t^2)^2}$$

and this is the direction of the normal. At $t = 1$, the unit tangent vector is $T = \frac{i+2j+2k}{3}$ and a normal vector is $\frac{-4i-2j+4k}{9}$. Therefore, the equation of the tangent is : $(x, y, z) = (1, 1, \frac{2}{3}) + t(1, 2, 2)$ and the equation of the principal normal is : $(x, y, z) = (1, 1, \frac{2}{3}) + t(-2, -1, 2)$. The direction of the binormal is defined to $b = T \times N$. Simple calculation will lead to the equation of the binormal.

To evaluate the curvature we can use the formula $\kappa(t) = \frac{\|T'(t)\|}{\|\frac{dR}{dt}\|}$ as we have already evaluated $T'(t)$ or the formula given in the previous theorem. In any way, we get $\kappa(t) = \frac{2}{(1+2t^2)^2}$.

Lectures 26-27: Functions of Several Variables
(Continuity, Differentiability, Increment Theorem and Chain Rule)

The rest of the course is devoted to calculus of several variables in which we study continuity, differentiability and integration of functions from $\mathbb{R}^n$ to $\mathbb{R}$, and their applications.

In calculus of single variable, we had seen that the concept of convergence of sequence played an important role, especially, in defining limit and continuity of a function, and deriving some properties of $\mathbb{R}$ and properties of continuous functions. This motivates us to start with the notion of convergence of a sequence in $\mathbb{R}^n$. For simplicity, we consider only $\mathbb{R}^2$ or $\mathbb{R}^3$. General case is entirely analogous.

Convergence of a sequence : Let $X_n = (x_{1,n}, x_{2,n}, x_{3,n}) \in \mathbb{R}^3$. We say that the sequence (X_n) is convergent if there exists $X_0 \in \mathbb{R}^3$ such that $\|X_n - X_0\| \to 0$ as $n \to \infty$. In this case we say that X_n converges to X_0 and we write $X_n \to X_0$.

Note that corresponding to a sequence (X_n), $X_n = (x_{1,n}, x_{2,n}, x_{3,n})$, there are three sequences $(x_{1,n})(x_{2,n})$ and $(x_{3,n})$ in $\mathbb{R}$, and vice-versa. Thus the properties of (X_n) can be completely understood in terms of the properties of the corresponding sequences $(x_{1,n})(x_{2,n})$ and $(x_{3,n})$ in $\mathbb{R}$. For example,

(i) $X_n \to X_0$ in $\mathbb{R}^3 \Leftrightarrow$ the coordinates $x_{i,n} \to x_{i,0}$ for every $i = 1, 2, 3$ in $\mathbb{R}$.

(ii) (X_n) is bounded (i.e., $\exists\ M$ such that $\|X_n\| \leq M\ \forall\ n$) $\Leftrightarrow$ each sequence $(x_{i,n}), i = 1, 2, 3,$ is bounded.

Using the previous idea, we can prove the following results.

Problem 1: *Every convergent sequence* $\mathbb{R}^3$ *is bounded.*

Problem 2 (Bolzano-Weierstrass Theorem): *Every bounded sequence in* $\mathbb{R}^3$ *has a convergent subsequence.*

In case of a sequence in $\mathbb{R}$, to define the notion of convergence or boundedness, we use $|\ \ |$ in place of $\|\ \ \|$, hence it is clear how we generalized the concept of convergence or boundedness of a sequence in $\mathbb{R}^1$ to $\mathbb{R}^3$. Moreover, it is also now clear how to define the concepts of limit and continuity of a function $f : \mathbb{R}^3 \to \mathbb{R}$ at some point $X_0 \in \mathbb{R}^3$.

Limit and Continuity : (i) We say that L is the limit of a function $f : \mathbb{R}^3 \to \mathbb{R}$ at $X_0 \in \mathbb{R}^3$ (and we write $\lim_{X \to X_0} f(X) = L$) if $f(X_n) \to L$ whenever a sequence (X_n) in $\mathbb{R}^3$, $X_n \neq X_0$, converges to X_0.

(ii) A function $f : \mathbb{R}^3 \to \mathbb{R}$ is continuous at $X_0 \in \mathbb{R}^3$ if $\lim_{X \to X_0} f(X) = f(X_0)$.

Examples 1: (i) Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(x, y) = \frac{\sin^2(x-y)}{|x|+|y|}$ when $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$. We will show that this function is continuous at (0,0). Note that

$$|\ f(x, y) - f(0, 0)\ | \leq \frac{|\ x - y\ |^2}{|\ x\ | + |\ y\ |} \ \leq |\ x\ | + |\ y\ | \ \ (\text{or } |\ x - y\ |\ )$$

Therefore, whenever a sequence $(x_n, y_n) \to (0, 0)$, i.e, $x_n \to 0$ and $y_n \to 0$, we have $f(x_n, y_n) \to f(0, 0)$. Hence f is continuous at $(0, 0)$. In fact, this function is continuous on the entire $\mathbb{R}^2$.

(ii) Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(x, y) = \frac{xy}{\sqrt{x^2+y^2}}$ when $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$. This function is continuous at $(0, 0)$, because, $|\ \frac{xy}{\sqrt{x^2+y^2}}| \leq \frac{|x^2+y^2|}{\sqrt{x^2+y^2}} = \sqrt{x^2 + y^2} \to 0$, as $(x, y) \to 0$.

(iii) Let $f(x,y) = \frac{2xy}{x^2+y^2}, (x,y) \neq (0,0)$. We will show that this function does not have a limit at $(0,0)$. Note that $f(x,mx) \to \frac{2m}{1+m^2}$ as $x \to 0$ for any m. This shows that the function does not have a limit at $(0,0)$.

(iv) Let $f(x,y) = \frac{x^2y}{x^4+y^2}$ when $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Note that $f(x,mx) \to 0$ as $x \to 0$. But the function is not continuous at $(0,0)$ because $f(x,x^2) \to \frac{1}{2}$ as $x \to 0$. Similarly we can show that the function $f(x,y)$ defined by $f(x,y) = \frac{x^4-y^2}{x^4+y^2}$ when $(x,y) \neq (0,0)$ and $f(0,0) = 0$ is not continuous at $(0,0)$ by taking $y = mx^2$ and allowing $x \to 0$.

Partial derivatives : The partial derivative of f with respect to the first variable at $X_0 = (x_0, y_0, z_0)$ is defined by

$$\frac{\partial f}{\partial x} |_{X_0} = \lim_{h \to 0} \frac{f(x_0 + h, y_0, z_0) - f(x_0, y_0, z_0)}{h}$$

provided the limit exists. Similarly we define $\frac{\partial f}{\partial y} |_{X_0}$ and $\frac{\partial f}{\partial z} |_{X_0}$.

Example 2: The function f defined by $f(x,y) = \frac{2xy}{x^2+y^2}$ *at* $(x,y) \neq (0,0)$ and $f(0,0) = 0$ is not continuous at $(0,0)$, however, the partial derivatives exist at $(0,0)$.

Problem 3: *Let $f(x,y)$ be defined in $S = \{(x,y) \in \mathbb{R}^2 : a < x < b,\ c < y < d\}$. Suppose that the partial derivatives of f exist and are bounded in S. Then show that f is continuous in S.*

Solution : Let $|f_x(x,y)| \leq M$ and $|f_y(x,y)| \leq M$ for all $(x,y) \in S$. Then

$$\begin{aligned} f(x+h, y+k) - f(x,y) &= f(x+h, y+k) - f(x+h, y) + f(x+h, y) - f(x,y) \\ &= kf_y(x+h, y+\theta_1 k) + hf_x(x+\theta_2 h, y), \text{ (for some } \theta_1, \theta_2 \in \mathbb{R}, \text{ by the MVT).} \end{aligned}$$

Hence, $|f(x+h, y+k) - f(x,y)| \leq M(|h| + |k|) \leq 2M\sqrt{h^2+k^2}$.

Hence, for $\epsilon > 0$, choose $\delta = \frac{\epsilon}{2M}$ or use the sequential argument to show that the function is continuous. □

It is clear from the previous example that the concept of differentiability of a function of several variables should be stronger than mere existence of partial derivatives of the function.

Differentiability : When $f : \mathbb{R} \to \mathbb{R}, x \in \mathbb{R}$ we define

$$f^{'}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \qquad (*)$$

provided the limit exists. In case $f : \mathbb{R}^3 \to \mathbb{R}$ and $X = (x_1, x_2, x_3) \in \mathbb{R}^3$ the above definition of the differentiability of functions of one variable (*) cannot be generalized as we cannot divide by an element of $\mathbb{R}^3$. So, in order to define the concept of differentiability, what we do is that we rearrange the above definition (*) to a form which can be generalized.

Let $f : \mathbb{R} \to \mathbb{R}$. Then f is differentiable at x if and only if there exists $\alpha \in \mathbb{R}$ such that

$$\frac{|\ f(x+h) - f(x) - \alpha \cdot h\ |}{|\ h\ |} \to 0 \quad as \quad h \to 0.$$

When f is differentiable at x, α has to be $f^{'}(x)$. We generalize this definition to the functions of several variables.

Definition : Let $f : \mathbb{R}^3 \to \mathbb{R}$ and $X = (x_1, x_2, x_3)$. We say that f is *differentiable* at X if there exists $\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3$ such that the error function

$$\varepsilon(H) = \frac{f(X+H) - f(X) - \alpha \cdot H}{\| H \|}$$

tends to 0 as $H \to 0$.

In the above definition $\alpha \cdot H$ is the scalar product. Note that the derivative $f'(X) = (\alpha_1, \alpha_2, \alpha_3)$.

Theorem 26.1: Let $f : \mathbb{R}^3 \to \mathbb{R}$, $X \in \mathbb{R}^3$. If f is differentiable at X then f is continuous at X.

Proof : Suppose f is differentiable at X. Then there exists $\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3$ such that

$$\mid f(X+H) - f(X) - \alpha \cdot H \mid = \| H \| \, \varepsilon(H) \text{ and } \varepsilon(H) \to 0 \text{ as } H \to 0.$$

Hence

$$\mid f(X+H) - f(X) \mid \le \| H \| \, (\sum_{i=1}^{3} \mid \alpha_i \mid) + \| H \| \, \varepsilon(H)$$

and $\varepsilon(H) \to 0$ as $H \to 0$. Therefore $f(X+H) \to f(X)$ as $H \to 0$. This proves that f is continuous at X. $\square$

How do we verify that a given function is differentiable at a point in $\mathbb{R}^3$? The following result helps us to answer this question.

Theorem 26.2: Suppose f is differentiable at X. Then the partial derivatives $\frac{\partial f}{\partial x} \mid_X$, $\frac{\partial f}{\partial y} \mid_X$ and $\frac{\partial f}{\partial z} \mid_X$ exist and the derivative

$$f^{'}(X) = (\alpha_1, \alpha_2, \alpha_3) = (\frac{\partial f}{\partial x} \mid_X, \frac{\partial f}{\partial y} \mid_X, \frac{\partial f}{\partial z} \mid_X).$$

Proof : Suppose f is differentiable at X and $f'(X) = (\alpha_1, \alpha_2, \alpha_3)$. Then by taking $H = (t, 0, 0)$, we have

$$\varepsilon(H) = \frac{f(X+H) - f(X) - \alpha_1 t}{|t|} \to 0 \;\; as \;\; t \to 0, \;\; i.e., \;\; \frac{f(X+H) - f(X) - \alpha_1 t}{t} \to 0$$

This implies that $\alpha_1 = \frac{\partial f}{\partial x} \mid_X$. Similarly we can show that $\alpha_2 = \frac{\partial f}{\partial y} \mid_X$ and $\alpha_3 = \frac{\partial f}{\partial x} \mid_X$. $\square$

Example 3 : Let

$$f(x, y) = xy \frac{x^2 - y^2}{x^2 + y^2} \;\; at \;\; (x, y) \neq (0, 0)$$
$$= 0 \;\; at \;\; (0, 0)$$

To verify that f is differentiable at $(0, 0)$, let us choose $\alpha = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \mid_{(0,0)}$ and verify that $\epsilon(H) \to 0$ as $H = (h, k) \to 0$. In this case $\alpha = (0, 0)$ and

$$\mid \varepsilon(H) \mid = \mid \frac{f(0+H) - f(0) - (0,0) \cdot H}{\| H \|} \mid \le \mid \frac{hk}{\sqrt{h^2 + k^2}} \mid \le \sqrt{h^2 + k^2} \to 0 \;\; as \;\; H \to 0.$$

Hence f is differentiable at $(0, 0)$. $\square$

Example 2 illustrates that the partial derivatives of a function at a point may exist but the function need not be differentiable at that point. The previous theorem says that if the function is differentiable at X then the derivative $f'(X)$ can be expressed in terms of the partial derivatives of f at X. Since finding partial derivatives is easy because they are based on one variable and it is related to the derivative, one naturally asks the following question: Under what additional assumptions on the partial derivatives the function becomes differentiable. The following criterion answer this question.

Theorem 26.3: If $f : \mathbb{R}^3 \to \mathbb{R}$ is such that all its partial derivatives exist in a neighborhood of X_0 and continuous at X_0 then f is differentiable at X_0.

We omit the proof of this result. We will see in a tutorial class that the converse of the previous result is not true.

Chain Rule: We have seen that the chain rule which deals with derivative of a function of a function is very useful in one variable calculus. In order to derive a similar rule for functions of several variables we need the following theorem called **Increment Theorem**. For simplicity we will state this theorem only for two variables.

We will employ the notation $f_x = \frac{\partial f}{\partial x}$ and $f_y = \frac{\partial f}{\partial x}$.

Theorem 26.4: Let $f(x, y)$ be differentiable at (x_0, y_0). Then we have

$$f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) \;=\; f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y$$

where $\varepsilon_1(\Delta x, \Delta y), \varepsilon_2(\Delta x, \Delta y) \to 0$ as $\Delta x \to 0$ and $\Delta y \to 0$.

Proof (*): Let $H = (\Delta x, \Delta y)$. Since the function is differentiable at (x_0, y_0), we have

$$f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) = f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y + \parallel H \parallel \varepsilon(H), \varepsilon(H) \to 0 \text{ as } H \to 0.$$

We have to show that $\parallel H \parallel \varepsilon(H) \;=\; \varepsilon_1 \Delta x + \varepsilon_2 \Delta y$ for some functions ε_1 and ε_2. Note that

$$\varepsilon(H) \parallel H \parallel = \;\frac{\varepsilon(H)}{\parallel H \parallel}(\Delta x^2 + \Delta y^2) \;=\; (\Delta x \frac{\varepsilon(H)}{\parallel H \parallel})\Delta x + (\Delta y \frac{\varepsilon(H)}{\parallel H \parallel})\Delta y.$$

Define $\varepsilon_1(H) = \Delta x \frac{\varepsilon(H)}{\|H\|}$ and $\varepsilon_2(H) = \Delta y \frac{\varepsilon(H)}{\|H\|}$. Note that

$$\mid \varepsilon_1(H) \mid = \mid \Delta x \frac{\varepsilon(H)}{\parallel H \parallel} \mid \leq \mid \varepsilon(H) \mid \to 0 \text{ as } \; H \to 0.$$

Similarly we can show that $\varepsilon_2(H) \to 0$ as $H \to 0$. This proves the result. $\square$

In the next result we present the chain rule.

Theorem 26.5: Let $f(x, y)$ be differentiable (or f has continuous partial derivatives) and if $x = x(t), y = y(t)$ are differentiable functions on t, then the function $w = f(x(t), y(t))$ is differentiable at t and

$$\frac{df}{dt} = f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t), \;\; \text{i.e.,} \;\; \frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Proof : By increment theorem we have

$$\Delta f \;=\; f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y, \quad \varepsilon_1, \varepsilon_2 \to 0 \;\text{ as }\; \Delta x, \Delta y \to 0$$

This implies that

$$\frac{\Delta f}{\Delta t} = f_x \frac{\Delta x}{\Delta t} + f_y \frac{\Delta y}{\Delta t} + \varepsilon_1 \frac{\Delta x}{\Delta t} + \varepsilon_2 \frac{\Delta y}{\Delta t}.$$

Allow $\Delta t \to 0$, which implies that $\varepsilon_1, \varepsilon_2 \to 0$ because $\Delta x, \Delta y \to 0$. Therefore, we get $\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$. $\square$

Lecture 28 : Directional Derivatives, Gradient, Tangent Plane

The partial derivative with respect to x at a point in $\mathbb{R}^3$ measures the rate of change of the function along the X-axis or say along the direction $(1,0,0)$. We will now see that this notion can be generalized to any direction in $\mathbb{R}^3$.

Directional Derivative : Let $f:\mathbb{R}^3 \to \mathbb{R}$, $X_0 \in \mathbb{R}^3$ and $U \in \mathbb{R}^3$ such that $\parallel U \parallel = 1$. The directional derivative of f in the direction U at $X_0 = (x_0, y_0, z_0)$ is defined by

$$D_{X_0} f(U) = \lim_{t \to 0} \frac{f(X_0 + tU) - f(X_0)}{t}$$

provided the limit exists.

It is clear that $D_{X_0} f(e_1) = f_x(X_0)$, $D_{X_0} f(e_2) = f_y(X_0)$ and $D_{X_0} f(e_3) = f_z(X_0)$.

The proof of the following theorem is similar to the proof of Theorem 26.2.

Theorem 28.1: *If f is differentiable at X_0, then $D_{X_0} f(U)$ exists for all $U \in \mathbb{R}^3$, $\parallel U \parallel = 1$. Moreover, $D_{X_0} f(U) = f'(X_0) \cdot U = (f_x(X_0), f_y(X_0), f_z(X_0)) \cdot U$.*

The previous theorem says that if a function is differentiable then all its directional derivatives exist and they can be easily computed from the derivative.

Examples :

(i) In this example we will see that a function is not differentiable at a point but the directional derivatives in all directions at that point exist.

Define $f:\mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = \frac{x^2 y}{x^4 + y^2}$ when $(x,y) \neq (0,0)$ and $f(0,0) = 0$.

This function is not continuous at $(0,0)$ and hence it is not differentiable at $(0,0)$.

We will show that the directional derivatives in all directions at $(0,0)$ exist. Let $U = (u_1, u_2) \in \mathbb{R}^3$, $\parallel U \parallel = 1$ and $\mathbf{0} = (0,0)$. Then

$$\lim_{t \to 0} \frac{f(\mathbf{0} + tU) - f(\mathbf{0})}{t} = \lim_{t \to 0} \frac{t^3 u_1^2 u_2}{t(t^4 u_1^4 + t^2 u_2^2)} = \lim_{t \to 0} \frac{u_1^2 u_2}{t^2 u_1^4 + u_2^2} = 0, \text{ if } u_2 = 0 \text{ and } \frac{u_1^2}{u_2}, \text{ if } u_2 \neq 0$$

Therefore, $D_{\mathbf{0}} f((u_1, 0)) = 0$ and $D_{\mathbf{0}} f((u_1, u_2)) = \frac{u_1^2}{u_2}$ when $u_2 \neq 0$.

(ii) In this example we will see that the directional derivative at a point with respect to some vector may exist and with respect to some other vector may not exist.

Consider the function $f(x,y) = \frac{x}{y}$ if $y \neq 0$ and 0 if $y = 0$. Let $U = (u_1, u_2)$ and $\parallel U \parallel = 1$. It is clear that if $u_1 = 0$ or $u_2 = 0$, then $D_{\mathbf{0}} f(U)$ exists and is equal to 0. If $u_1 u_2 \neq 0$ then

$$\lim_{t \to 0} \frac{f(\mathbf{0} + tU) - f(\mathbf{0})}{t} = \lim_{t \to 0} \frac{u_1}{t u_2}$$

does not exist. So, only the partial derivatives of the function at $\mathbf{0}$ exist. Note that this function can not be differentiable at $\mathbf{0}$ (Why ?).

Problem 1: *Let $f(x,y) = \frac{y}{|y|}\sqrt{x^2 + y^2}$ if $y \neq 0$ and $f(x,y) = 0$ if $y = 0$. Show that f is continuous at $(0,0)$, it has all directional derivatives at $(0,0)$ but it is not differentiable at $(0,0)$.*

Solution : Note that $\mid f(x,y) - f(0,0) \mid = \sqrt{x^2 + y^2}$. Hence the function is continuous.

For $\| (u_1, u_2) \| = 1$, $\lim_{t\to 0} \frac{f(tu_1,tu_2)}{t} = 0$ if $u_2 = 0$ and $\frac{u_2}{|u_2|}$ if $u_2 \neq 0$. Therefore directional derivatives in all directions exist.

Note that $f_x(0,0) = 0$ and $f_y(0,0) = 1$. If f is differentiable at $(0,0)$ then $f'(0,0) = \alpha = (0,1)$. Note that

$$\epsilon(h,k) = \frac{\frac{k}{|k|}\sqrt{h^2+k^2} - k}{\sqrt{h^2+k^2}} \nrightarrow 0 \quad \text{as} \quad (h,k) \to (0,0).$$

For example, $h = k$ gives $(\sqrt{2}-1)\frac{k}{|k|} \nrightarrow 0$ as $k \to 0$. Therefore the function is not differentiable at $(0,0)$. $\square$

The vector $(f_x(X_0), f_y(X_0), f_z(X_0))$ is called **gradient** of f at X_0 and is denoted by $\nabla f(X_0)$.

An Application : Let us see an application of Theorem 1. Suppose f is differentiable at X_0. Then $f'(X_0) = \nabla f(X_0)$ and $D_{X_0}f(U) = \nabla f(X_0) \cdot U = \| \nabla f(X_0) \| \cos\theta$ where $\theta \in [0,\pi]$ is the angle between the gradient and U. Suppose $\nabla f(X_0) \neq 0$. Then $D_{X_0}f(U)$ is maximum when $\theta = 0$ and minimum $\theta = \pi$. That is, f increases (respectively, decreases) most rapidly around X_0 in the direction $U = \frac{\nabla f(X_0)}{\|\nabla f(X_0)\|}$ (respectively, U = -$\frac{\nabla f(X_0)}{\|\nabla f(X_0)\|}$).

Example: Suppose the temperature of a metallic sheet is given as $f(x,y) = 20 - 4x^2 - y^2$. We will start from the point $(2,1)$ and find a path i.e., a plane curve, $r(t) = x(t)i + y(t)j$ which is a path of maximum increase in the temperature. Note that the direction of the path is $r'(t)$. This direction should coincide with that of the maximum increase of f. Therefore, $\alpha r'(t) = \nabla f$ for some α. This implies that $\alpha x'(t) = -8x$ and $\alpha y'(t) = -2y$. By chain rule we have $\frac{dy}{dx} = \frac{2y}{8x} = \frac{y}{4x}$. Since the curve passes through (2,1), we get $x = 2y^4$.

We will now see a geometric interpretation of the derivative i.e, gradient.

Tangent Plane: Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is differentiable and $c \in \mathbb{R}$. Consider the surface $S = \{(x,y,z) : f(x,y,z) = c\}$. This surface is called a level surface at the height c. (For example if $f(x,y,z) = x^2 + y^2 + z^2$ and $c = 1$, then S is the unit sphere.) Let $P = (x_0, y_0, z_0)$ be a point on S and $R(t) = (x(t), y(t), z(t))$ be a differentiable (i.e., smooth) curve lying on S. With these assumptions we prove the following result.

Theorem 28.2: *If T is the tangent vector to $R(t)$ at P then $\nabla f(P) \cdot T = 0$.*

Proof : Since $R(t)$ lies on S, $f(x(t), y(t), z(t)) = c$. Hence $\frac{df}{dt} = 0$. By chain rule,

$$\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = 0 \quad i.e., \ \nabla f \cdot \frac{dR}{dt} = 0 \quad i.e., \ \nabla f \cdot T = 0 \text{ at } P. \qquad \square$$

From the previous theorem we conclude the following. Note that the gradient $\nabla f(P)$ is perpendicular to the tangent vector to every smooth curve $R(t)$ on S passing through P. That is, all these tangent vectors lie on a plane which is perpendicular to $\nabla f(P)$. That is, $\nabla f(P)$, when $\nabla f(P) \neq 0$, is the normal to the surface at P. Therefore, the plane through P with normal $\nabla f(P)$ defined by

$$f_x(P)(x - x_0) + f_y(P)(y - y_0) + f_z(P)(z - z_0) = 0$$

is called the tangent plane to the surface S at $P = (x_0, y_0, z_0)$.

Suppose the surface is given as a graph of $f(x,y)$, i.e., $S = \{(x, y, f(x,y)) : (x,y) \in D \subseteq \mathbb{R}^2\}$. Then it can be considered as a level surface $S = \{(x,y,z) : F(x,y,z) = 0\}$ where $F(x,y,z) = f(x,y) - z$. Let $X_0 = (x_0, y_0)$, $z_0 = f(x_0, y_0)$ and $P = (x_0, y_0, z_0)$. Then the equation of the tangent plane is $f_x(x_0,y_0)(x - x_0) + f_y(x_0,y_0)(y - y_0) - (z - z_0) = 0$ i.e.,

$$z = f(X_0) + f'(X_0)(X - X_0), \quad X = (x,y) \in \mathbb{R}^2.$$

Lecture 29 : Mixed Derivative Theorem, MVT and Extended MVT

If $f : \mathbb{R}^2 \to \mathbb{R}$, then f_x is a function from $\mathbb{R}^2$ to $\mathbb{R}$ (if it exists). So one can analyze the existence of

$$f_{xx} = (f_x)_x = \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(\frac{\partial f}{\partial x}) \quad \text{and} \quad f_{xy} = (f_x)_y = \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(\frac{\partial f}{\partial x})$$

which are partial derivatives of f_x with respect x or y and, similarly the existence of f_{yy} and f_{yx}. These are called second order partial derivatives of f.

The following example shows that, in general, f_{xy} need not be equal to f_{yx}.

Example : Let $f(x,y) = xy\frac{x^2-y^2}{x^2+y^2}$ if $(x,y) \neq (0,0)$ and $f(0,0) = 0$. Note that

$$f_y(h,0) = \lim_{k\to 0} \frac{f(h,k) - f(h,0)}{k} = \lim_{k\to 0} \frac{h(h^2-k^2)}{h^2+k^2} = h$$

and

$$f_{yx}(0,0) = \lim_{h\to 0} \frac{f_y(h,0) - f_y(0,0)}{h} = 1.$$

Similarly, $f_{xy}(0,0) = -1$.

Theorem 29.1(Mixed derivative theorem) : *If $f(x,y)$ and its partial derivatives f_x, f_y, f_{xy} and f_{yx} are defined in a neighborhood of (x_0,y_0) and all are continuous at (x_0,y_0) then $f_{xy}(x_0,y_0) = f_{yx}(x_0,y_0)$.*

We will not present the proof of this result here. The proof is given in the text book.

Mean Value Theorem : We will present the MVT for functions of several variables which is a consequence of MVT for functions of one variable.

Theorem 29.2: *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable. Let $X_0 = (x_0,y_0)$ and $X = (x_0+h, y_0+k)$. Then there exists C which lies on the line joining X_0 and X such that*

$$f(X) = f(X_0) + f'(C)(X - X_0)$$

i.e, there exists $c \in (0,1)$ such that

$$f(x_0+h, y_0+k) = f(x_0,y_0) + hf_x(C) + kf_y(C) \ \textit{where} \ \ C = (x_0+ch, y_0+ck).$$

Proof : Define $\phi : [0,1] \to \mathbb{R}$ by

$$\phi(t) = f(x_0+th, y_0+tk), \ t \in [0,1].$$

Note that by the Chain Rule ϕ is differentiable and

$$\phi' = f_x\frac{dx}{dt} + f_y\frac{dy}{dt} = hf_x + kf_y.$$

By the MVT, there exist $c \in (0,1)$ such that

$$\phi(1) - \phi(0) = \phi'(c).$$

This proves the result. $\square$

Remark : In the previous result if we fix X_0 and X then it is enough to assume that the function f is differentiable on the line segment joining X and X_0.

Problem : *If $f(x,y)$ is constant if and only if $f_x = 0$ and $f_y = 0$.*

We will now take up the extended mean value theorem which we need.

Theorem 29.3(EMVT): *Let f, X, X_0 be as in the previous theorem. Suppose f_x and f_y are continuous and they have continuous partial derivatives. Then, there exists C which lies on the line joining X_0 and X such that*

$$f(X) = f(X_0) + f'(X_0)(X - X_0) + \frac{1}{2}(X - X_0)f''(C)(X - X_0)$$

where $f'' = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$. That is, there exists $c \in (0,1)$ such that

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + (hf_x + kf_y)(X_0) + \frac{1}{2}(h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy})(C)$$

where $C = (x_0 + ch, y_0 + ck)$.

Proof (*): Consider the function $\phi(t)$ defined in the proof of the previous result. Since f_x and f_y are continuous f is differentiable. Therefore, as given in the proof of the previous theorem, ϕ is differentiable and

$$\phi' = f_x \frac{dx}{dt} + f_y \frac{dy}{dt} = hf_x + kf_y.$$

Since f_x and f_y have continuous partial derivatives, they are differentiable. Denote

$$\phi'(t) = hf_x(x_0 + th, y_0 + tk) + kf_y(x_0 + th, y_0 + tk) = F(x_0 + th, y_0 + tk),\ t \in [0,1].$$

Again by the Chain Rule,

$$\phi'' = hF_x + kF_y = h\frac{\partial}{\partial x}(hf_x + kf_y) + k\frac{\partial}{\partial y}(hf_x + kf_y) = h(h\frac{\partial^2 f}{\partial x^2} + k\frac{\partial^2 f}{\partial y \partial x}) + k(h\frac{\partial^2 f}{\partial x \partial y} + k\frac{\partial^2 f}{\partial y^2}).$$

By the mixed derivative theorem,

$$\phi'' = h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy}.$$

By the EMVT for ϕ, there exists $c \in (0,1)$ such that

$$\phi(1) = \phi(0) + \phi'(0) + \frac{\phi''(c)}{2}.$$

By substituting ϕ, ϕ' and ϕ'' in the above equation we get the result. $\square$

Remarks : 1. We will consider f'', given in the statement of the previous theorem, as a notation. We do not say that the function f is twice differentiable.

2. We will recall the EMVT when we will deal with the second derivative test for local maxima and minima of $f(x,y)$ in the next lecture.

3. Whatever we discussed above can be generalized to the functions of three variables.

4. The matrix given in the statement of the previous theorem is called Hessian matrix. We should be able to guess what should be the corresponding Hessian matrix for the functions of three variables.

5. Note that we applied the MVT and the EMVT for the function ϕ to get the MVT and the EMVT for $f(x,y)$. Similarly by assuming that $f(x,y)$ has continuous partial derivatives of order n and applying Taylor's theorem for the function ϕ, we can obtain Taylor's Theorem for $f(x,y)$.

Lecture 30 : Maxima, Minima, Second Derivative Test

In calculus of single variable we applied the Bolzano-Weierstrass theorem to prove the existence of maxima and minima of a continuous function on a closed bounded interval. Moreover, we developed first and second derivative tests for local maxima and minima. In this lecture we will see a similar theory for functions of several variables.

Definition : A non-empty subset D of $\mathbb{R}^n$ is said to be closed if a sequence in D converges then its limit point lies in D.

For example, the sets $B_1 = \{X_0 \in \mathbb{R}^2 : \| X \| \leq 1\}$ and $H = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0\}$ are closed subsets of $\mathbb{R}^2$. However, the sets $S_1 = \{X \in \mathbb{R}^2 : \| X \| < 1\}$ and $H^+ = \{(x, y) \in \mathbb{R}^2 : x > 0, y > 0\}$ are not closed.

Definition : Let $D \subseteq \mathbb{R}^n$ and $X_0 \in D$. We say that X_0 is an interior point of D if there exists $r > 0$ such that the neighborhood $N_r(X_0) = \{X \in \mathbb{R}^n : \| X_0 - X \| < r\}$ is contained in D.

For example, all the points of S_1 are interior points of B_1. Similarly, all the points of H^+ are interior points of H.

The notions of maxima, minima, local maxima and local minima are similar to the ones defined for the functions of one variable. The proof of the following theorem is similar to the proof of the existence of maximum and minimum of a continuous function on a closed bounded interval.

Theorem 30.1(Existence of Maxima and Minima): *Let D be a closed and bounded subset of $\mathbb{R}^2$ and $f : D \to \mathbb{R}$ be continuous. Then f has a maximum and a minimum in D.*

Theorem 30.2(Necessary Condition for Local Maximum and Minimum): *Suppose $D \subseteq \mathbb{R}^2$, $f : D \to \mathbb{R}$ and (x_0, y_0) is an interior point of D. Let f_x and f_y exist at the point (x_0, y_0). If f has a local maximum or local minimum at (x_0, y_0) then $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$.*

Proof : Note that (the one variable) functions $f(x, y_0)$ and $f(x_0, y)$ have local maximum or minimum at x_0 and y_0 respectively. Therefore, the derivatives of these functions are zero at x_0 and y_0 respectively. That is, $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$. $\square$

Note that the conditions given in the previous results are not sufficient. For example, consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = xy$. Note that $f_x(0, 0) = f_y(0, 0) = 0$ but $(0, 0)$ is neither a local minimum nor a local maximum for f.

Second Derivative Test for Local Maximum and Local Minimum : Suppose $D \subseteq \mathbb{R}^2$ and $f : D \to \mathbb{R}$. Suppose f_x and f_y are continuous and they have continuous partial derivatives on D. With these assumptions we prove the following result.

Theorem 30.3: *Let (x_0, y_0) be an interior point of D and $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$. Suppose $(f_{xx}f_{yy} - f_{xy}^2)(x_0, y_0) > 0$. Then*

(i) if $f_{xx}(x_0, y_0) > 0$ then f has a local minimum at (x_0, y_0).

(ii) if $f_{xx}(x_0, y_0) < 0$ then f has a local maximum at (x_0, y_0).

Proof (*): We prove (i) and the proof of (ii) is similar. Suppose $f_{xx}(x_0, y_0) > 0$ and $(f_{xx}f_{yy} - f_{xy}^2)(x_0, y_0) > 0$. Then there exists a neighborhood N of (x_0, y_0), such that

$$f_{xx}(x, y) > 0 \text{ and } (f_{xx}f_{yy} - f_{xy}^2)(x, y) > 0 \quad \text{for all} \quad (x, y) \in N.$$

Let $(x_0 + h, y_0 + k) \in N$. Them by the Extended MVT (applying over N, which is possible), there exists some C lying in the line joining (x_0+h, y_0+k) and (x_0, y_0) such that

$$f(x_0+h, y_0+k) - f(x_0, y_0) = Q(C) = \frac{1}{2}(h^2 f_{xx} + 2hk f_{xy} + k^2 f_{yy})(C).$$

Note that $2f_{xx}(C)Q(C) = \{(hf_{xx} + kf_{xy})(C)\}^2 + k^2(f_{xx}f_{yy} - f_{xy}^2)(C) \ > \ 0.$
Since $f_{xx}(C) > 0$ we have $Q(C) > 0$ and hence $f(x_0+h, y_0+k) > f(x_0, y_0)$. Therefore, f has a local minimum at (x_0, y_0). $\square$

Remarks : 1. If (x_0, y_0) is an interior point of D, $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ and $(f_{xx}f_{yy} - f_{xy}^2)(x_0, y_0) < 0$, then one can show that in every neighborhood of (x_0, y_0) we can find two points (x_1, y_1) and (x_2, y_2) such that $f(x_1, y_1) > f(x_0, y_0)$ and $f(x_2, y_2) < f(x_0, y_0)$, that is (x_0, y_0) is a saddle point.

2. The above test is inconclusive when $f_x(x_0, y_0) = f_y(x_0, y_0) = (f_{xx}f_{yy} - f_{xy}^2)(x_0, y_0) = 0$.

Examples : 1. The functions $f_1(x, y) = -(x^4 + y^4)$ and $f_2(x, y) = x^4 + y^4$ satisfy the above equation for $(x_0, y_0) = (0, 0)$ but f_1 has a local maximum at $(0, 0)$ and f_2 has a local minimum at $(0, 0)$.

2. Consider the function $f(x, y) = (x+y)^2 - x^4$. This function satisfies the above equation for $(x_0, y_0) = (0, 0)$ but it has neither a local maximum nor a local minimum at $(0, 0)$. In fact, $(0, 0)$ is a saddle point. This can be verified as follows. Note that for $0 < x < 1$, $f(x, x) > 0$ and $f(x, -x) < 0$.

3. Let $f(x, y) = x \sin y$. Here $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ for $(x_0, y_0) = (0, n\pi), n \in \mathbb{N}$. Note that $(f_{xx}f_{yy} - f_{xy}^2)(x_0, y_0) < 0$. Therefore, the points $(0, n\pi), n \in \mathbb{N}$ are saddle points.

Problem 1: Let $f(x, y) = 3x^4 - 4x^2y + y^2$. Show that f has a local minimum at $(0, 0)$ along every line through $(0, 0)$. Does f have a minimum at $(0, 0)$? Is $(0, 0)$ a saddle point for f?

Solution : Let $f(x, y) = 3x^4 - 4x^2y + y^2$. Along, the x-axis, the local minimum of the function is at $(0, 0)$. Let $x = r\cos\theta$ and $y = r\sin\theta$, for a fixed $\theta \neq 0, \pi$ (or let $y = mx$). Then, $f(r\cos\theta, r\sin\theta) = 3r^4\sin^4\theta - 4r^3\cos^2\theta\sin\theta + r^2\sin^2\theta$ which is a function of one variable. By the second derivative test (for functions of one variable), we see that $(0, 0)$ is a local minima. Since, $f(x, y) = (3x^2 - y)(x^2 - y)$, we see that in the region between the parabolas $3x^2 = y$ and $y = x^2$, the function takes negative values and is positive everywhere else. Thus, $(0, 0)$ is a saddle point for f.

Problem 2: Let $D = [-2, 2] \times [-2, 2]$ and $f : D \to \mathbb{R}$ be defined as $f(x, y) = 4xy - 2x^2 - y^4$. Find absolute maxima and absolute minima of f in D.

Solution (Hints) : Note that $f_x(x_0, y_0) = f_y(x_0, y_0) = 0$ for $(x_0, y_0) = (0, 0), (1, 1)$ or $(-1, -1)$. Since these points lie in the interior of D, these are the candidates for maxima and minima for f on the set of interiors of D.

Now we have to check the behavior of the function over the boundary of D. Note that $(x, y) \in D$ is a boundary point if and only if $x = \pm 2$ or $y = \pm 2$. So we have to consider the functions $f(2, y), f(-2, y), f(x, 2)$ and $f(x, -2)$ over the interval $[-2, 2]$. For example, $f(2, y) = 8y - 8 - y^4, y \in [-2, 2]$, has absolute maximum at $y = \sqrt[3]{2}$ and absolute minimum at $y = -2$. So, $(2, \sqrt[3]{2})$ and $(2, -2)$ are the candidates for maxima and minima on the boundary line $\{(2, y) : y \in [-2, 2]\}$. Find all possible candidates for maxima and minima and choose the maxima and minima from these candidates.

The absolute maximum value of f on D is 1 which is obtained at $(1, 1)$ and $(-1, -1)$. The absolute minimum value of f on D is -40 and which is obtained at $(2, -2)$ and $(-2, 2)$.

Lecture 31 : Lagrange Multiplier Method

Let $f : S \to \mathbb{R}$, $S \subset \mathbb{R}^3$ and $X_0 \in S$. If X_0 is an interior point of the constrained set S, then we can use the necessary and sufficient conditions (first and second derivative tests) studied in the previous lecture in order to determine whether the point is a local maximum or minimum (i.e., local extremum) of f on S. If X_0 is not an interior point then one cannot apply these tests. For example, one cannot apply these tests at a point on a sphere $x^2 + y^2 + z^2 = c^2$, because no point is an interior point in this constrained set.

In general, constrained extremum problems are very difficult to solve and there is no general method for solving such problems. In case the constrained set is a level surface, for example a sphere, there is a special method called Lagrange multiplier method for solving such problems. So, we will be dealing with the following type of problem.

Problem : Find the local or absolute maxima and minima of a function $f(x, y, z)$ on the (level) surface $S := \{(x, y, z) : g(x, y, z) = 0\}$ where $g : \mathbb{R}^3 \to \mathbb{R}$.

Let us illustrate the problem with an example.

Example : Find a point on the plane $\{(x, y, z) : 2x + 3y - z = 5\}$ which is nearest to the origin of $\mathbb{R}^3$. Note that here we are minimizing the function $f(x, y, z) = \sqrt{x^2 + y^2 + z^2}$ or $x^2 + y^2 + z^2$ over the constrained set S defined by $g(x, y, z) = 2x + 3y - z - 5$.

Necessary condition: Let $P_0 = (x_0, y_0, z_0)$ be a point on $S := \{(x, y, z) : g(x, y, z) = 0\}$. Suppose P_0 is a local extremum of f over S. Let us try to find a necessary condition. Our argument is going to be geometric and so we will not question certain assumptions made in the following argument.

Let us first assume that f and g have continuous partial derivatives and $\nabla g|_{P_0} \neq 0$. Let $C = R(t) = x(t)i + y(t)j + z(t)k$ be a curve on S passing through P_0 and let $P_0 = R(t_0)$. Since f has a local extremum at P_0 on S, it has a local extremum on C as well. Therefore $\frac{df}{dt}|_{t_0} = 0$. By the chain rule, $\nabla f \cdot \frac{dR}{dt} = 0$. Since the curve C is arbitrary, we conclude that $\nabla f|_{P_0}$ is perpendicular to the tangent plane of S at P_0. But we already know that $\nabla g|_{P_0}$ is also perpendicular to the tangent plane of S at P_0. Therefore, there exists $\lambda \in \mathbb{R}$ such that $\nabla f|_{P_0} = \lambda \nabla g|_{P_0}$.

So the following method is anticipated.

Lagrange Multiplier Method: Suppose f and g have continuous partial derivatives. Let $(x_0, y_0, z_0) \in S := \{(x, y, z) : g(x, y, z) = 0\}$ and $\nabla g(x_0, y_0, z_0) \neq 0$. If f has a local maximum or minimum at (x_0, y_0, z_0) then there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0).$$

To find the extremum points, in practice, we consider the following equations

$$\nabla f(x, y, z) = \lambda \nabla g(x, y, z) \quad \text{and} \quad g(x, y, z) = 0. \tag{1}$$

These equations are solved for the unknowns x, y, z and λ. Then the local extremum points are found among the solutions of these equations.

Let us illustrate the method with a few examples.

Examples: 1. Let us find a point on the plane $2x + 3y - z = 5$ in $\mathbb{R}^3$ which is nearest to the origin. We have to minimize the function $f(x, y, z) = x^2 + y^2 + z^2$ subject to the constraint $g(x, y, z) = 2x + 3y - z - 5 = 0$. Here note that $\nabla g \neq 0$ at all points. The equations given in (1) imply that

$$2x = 2\lambda, \; 2y = 3\lambda, \; 2z = -\lambda \text{ and } 2x + 3y - z - 5 = 0.$$

Substituting $x = \lambda$, $y = \frac{3\lambda}{2}$ and $z = -\frac{\lambda}{2}$ in the equation $2x + 3y - z - 5 = 0$, we obtain that $\lambda = \frac{5}{7}$ and hence $\lambda = \frac{5}{7}$ and $(x, y, z) = (\frac{5}{7}, \frac{15}{14}, -\frac{5}{14})$ satisfy the equation (1). Since f attains its minimum on the plane, by the Lagrange multipliers method, the point $(\frac{5}{7}, \frac{15}{14}, -\frac{5}{14})$ has to be the nearest point.

2. Consider the problem of minimizing the function $f(x, y) = x^2 + y^2$ subject to the condition $g(x, y) = (x - 1)^3 - y^2 = 0$. The problem is to find a point on the curve $y^2 = (x - 1)^3$ which is nearest to the origin of $\mathbb{R}^2$. Geometrically, it is clear that the point $(1, 0)$ is the nearest point. But $\nabla g(1, 0) = 0$ while $\nabla f(1, 0) = (2, 0)$. Therefore $\nabla f(1, 0) \neq \lambda \nabla g(1, 0)$ for any λ. This explains that the condition $\nabla g(x_0, y_0, z_0) \neq 0$ cannot be dropped from the Lagrange multiplier method and a point at which ∇g is $(0, 0)$ could be an extremum point.

3. Let us evaluate the minimum and maximum value of the function $f(x, y) = 2 - x^2 - 2y^2$ subject to the condition $g(x, y) = x^2 + y^2 - 1$. If we use the Lagrange multiplier method, the equations in (1) imply that $2x + 2\lambda x = 0$, $4y + 2\lambda y = 0$ and $x^2 + y^2 - 1 = 0$. From the first two equations, we must have either $\lambda = -1$ or $\lambda = -2$. If $\lambda = -1$, then $y = 0$, $x = \pm 1$ and $f(x, y) = 1$. Similarly, if $\lambda = -2$, then $y = \pm 1$, $x = 0$ and $f(x, y) = 0$. Since the continuous function $f(x, y)$ achieves its maximum and minimum over the closed and bounded set $x^2 + y^2 = 1$, the points $(0, \pm 1)$ are the minima and $(\pm 1, 0)$ are the maxima, and the maximum value and the minimum value of f are 1 and 0 respectively.

This problem can also be solved using substitution as follows. Since $x^2 + y^2 - 1 = 0$, $x^2 = 1 - y^2$. Substituting this into f, we get $f(x, y) = 1 - y^2$. We are back to a one variable problem, which has a maximum at $y = 0$, where $x = \pm 1$ and $f(x, y) = 1$. Since $y \in [-1, 1]$, f has a minimum at $y = \pm 1$ where $x = 0$ and $f(x, y) = 0$. Although we solved this problem easily using substitution, it is usually very hard to solve such constrained problems using substitution.

4. Given n positive numbers $a_1, a_2, .., a_n$, let us find the maximum value of the expression $a_1 x_1 + a_2 x_2 + ... + a_n x_n$ where $x_1^2 + ... + x_n^2 = 1$. Note that here $f(x_1, x_2, .., x_n) = a_1 x_1 + a_2 x_2 + ... + a_n x_n$ and $g(x_1, x_2, ..., x_n) = x_1^2 + ... + x_n^2 - 1$. Although we stated the Lagrange multiplier method in $\mathbb{R}^3$, it works in $\mathbb{R}^n$ as well. The equations in (1) imply that $a_1 = 2\lambda x_1, ..., a_n = 2\lambda x_n$ and $x_1^2 + ... + x_n^2 - 1 = 0$. Therefore, $a_1^2 + a_2^2 + .. + a_n^2 = 4\lambda^2$. This implies that $\lambda = \pm \frac{\sqrt{a_1^2 + .. + a_n^2}}{2}$. Since the continuous function f achieves its minimum and maximum on the closed and bounded subset $x_1^2 + ... + x_n^2 = 1$, $\lambda = \frac{\sqrt{a_1^2 + .. + a_n^2}}{2}$ leads to the maximum value $f(\frac{a_1}{2\lambda}, \frac{a_2}{2\lambda}, ..., \frac{a_n}{2\lambda}) = \sqrt{a_1^2 + .. + a_n^2}$ and $\lambda = -\frac{\sqrt{a_1^2 + .. + a_n^2}}{2}$ leads to the minimum value of f.

Problem 1: Assume that among all rectangular boxes with fixed surface area of 20 square meters, there is a box of largest possible volume. Find its dimensions.

Solution: Let the box have sides of length x, y, $z > 0$. Then $V(x, y, z) = xyz$ and $xy + yz + xz = 10$. Using the method of Lagrange multipliers, we see that $yz = \lambda(y+z)$, $xz = \lambda(x+z)$ and $xy = \lambda(x+y)$. It is easy to see that $x, y, z > 0$. Now, we can see that $x = y = z$ and therefore, $x = y = z = \sqrt{\frac{10}{3}}$.

Problem 2: A company produces steel boxes at three different plants in amounts x, y and z, respectively, producing an annual revenue of $f(x, y, z) = 8xyz^2 - 200(x + y + z)$. The company is to produce 100 units annually. How should the production be distributed to maximize revenue?

Solution: Here, $g(x, y, z) = x + y + z - 100$. The Lagrange multiplier method implies that $8yz^2 - 200 = \lambda, 8xz^2 - 200 = \lambda, 16xyz - 200 = \lambda$ and $x + y + z - 100 = 0$. These imply that $x = y, z = 2x$ and $x = 25$.

Lecture 32 : Double integrals

In one variable calculus we had seen that the integral of a nonnegative function is the area under the graph. The double integral of a nonnegative function $f(x, y)$ defined on a region in the plane is associated with the volume of the region under the graph of $f(x, y)$.

The definition of double integral is similar to the definition of Riemannn integral of a single variable function. Let $Q = [a, b] \times [c, d]$ and $f : Q \to \mathbb{R}$ be bounded. Let P_1 and P_2 be partitions of $[a, b]$ and $[c, d]$ respectively. Suppose $P_1 = \{x_0, x_1, ..., x_n\}$ and $P_2 = \{y_0, y_1, ..., y_n\}$. Note that the partition $P = P_1 \times P_2$ decomposes Q into mn sub-rectangles. Define $m_{ij} = \inf\{f(x, y) : (x, y) \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]\}$ and $L(P, f) = \sum_{i=1}^{n} \sum_{j=1}^{m} m_{ij} \triangle y_j \triangle x_i$. Similarly we can define $U(P, f)$. Define lower integral and upper integral as we do in the single variable case. We say that $f(x, y)$ is integrable if both lower and upper integral of $f(x, y)$ are equal. If the function $f(x, y)$ is integrable on Q then the double integral is denoted by

$$\iint\limits_{Q} f(x, y)dxdy \quad \text{or} \quad \iint\limits_{Q} f(x, y)dA.$$

The proof of the following theorem is similar to the single variable case.

Theorem: If a function $f(x, y)$ is continuous on a rectangle $Q = [a, b] \times [c, d]$ then f is integrable on Q.

Fubini's Theorem: In one variable case, we use the second FTC for calculating integrals. The following result, called Fubini's theorem, provides a method for calculating double integrals. Basically, it converts a double integral into two successive one dimensional integrations.

Theorem 32.1: Let $f : Q = [a, b] \times [c, d] \to \mathbb{R}$ be continuous. Then

$$\iint\limits_{Q} f(x, y)dxdy = \int_c^d \left(\int_a^b f(x, y)dx \right) dy = \int_a^b \left(\int_c^d f(x, y)dy \right) dx.$$

We will not present the proof of the previous theorem, instead we present a geometric interpretation of it.

Geometric interpretation: Let $f(x, y) > 0$ for every $(x, y) \in Q$ and f be continuous. Consider the solid S enclosed by Q, the planes $x = a,\ x = b,\ y = c,\ y = d$ and the surface $z = f(x, y)$. From the way we have defined the double integral, we can consider the value $\iint\limits_{Q} f(x, y)dxdy$ as the volume of S. We will now use the method of slicing and calculate the volume of S.

For every $y \in [c, d]$, $A(y) = \int_a^b f(x, y)dx$ is the area of the cross section of the solid S cut by a plane parallel to the xz-plane. Therefore, it follows from the method of slicing that

$$\int_c^d \left(\int_a^b f(x, y)dx \right) dy = \int_c^d A(y)dy$$

is the volume of the solid S. The other two successive single integrals compute the volume of S by integrating the area of the cross section cut by the planes parallel to the yz-plane.

Double integral over general bounded regions: We defined the double integral of a function which is defined over a rectangle. We will now extend the concept to more general bounded regions.

Let $fx, y)$ be a bounded function defined on a bounded region D in the plane. Let Q be a rectangle such that $D \subseteq Q$. Define a new function $\widetilde{f}(x, y)$ on Q as follows:

$$\widetilde{f}(x, y) = f(x, y) \text{ if } (x, y) \in D \quad \text{and} \quad \widetilde{f}(x, y) = 0 \text{ if } (x, y) \in Q \setminus D.$$

Basically we have extended the definition of f to Q by making the function value equal to 0 outside D. If $\widetilde{f}(x,y)$ is integrable over Q, then we say that $f(x,y)$ is integrable over D and we define $\iint\limits_D f(x,y)dxdy = \iint\limits_Q \widetilde{f}(x,y)dxdy$. We find that defining the concept of double integral over a more general region D is a trivial one, but the important question is how to evaluate $\iint\limits_D f(x,y)$.

If D is a general bounded domain, then there is no general method to evaluate the double integral. However, if the domain is in a simpler form (as given in the following result) then there is a result to convert the double integral in to two successive single integrals.

Fubini's theorem (stronger form) : Let $f(x,y)$ be a bounded function over a region D.

1. If $D = \{(x,y) : a \leq x \leq b \ \text{ and } \ f_1(x) \leq y \leq f_2(x)\}$ for some continuous functions $f_1, f_2 : [a,b] \to \mathbb{R}$, then $\iint\limits_D f(x,y)dxdy = \int\limits_a^b \left(\int\limits_{f_1(x)}^{f_2(x)} f(x,y)dy \right) dx.$
2. If $D = \{(x,y) : c \leq y \leq d \ \text{ and } \ g_1(y) \leq x \leq g_2(y)\}$ for some continuous functions $g_1, g_2 : [c,d] \to \mathbb{R}$, then $\iint\limits_D f(x,y)dxdy = \int\limits_c^d \left(\int\limits_{g_1(y)}^{g_2(y)} f(x,y)dx \right) dy.$

If we use the method of slicing, as we did earlier, we can get the geometric interpretation of the previous theorem. Let us illustrate the method given in the previous theorem with some examples.

Example 1: Let us evaluate the integral $\iint\limits_D (x+y)^2 dxdy$ where D is the region bounded by the lines joining the points $(0,0)$, $(0,1)$ and $(2,2)$. Note that the domain D (see Figure 1) is the form given in the first part of the previous theorem with $a = 0, b = 2, f_1(x) = x$ and $f_2(x) = \frac{x}{2} + 1$. Therefore, by the previous theorem $\iint\limits_D (x+y)^2 dxdy = \int\limits_0^2 \left(\int\limits_x^{\frac{x}{2}+1} (x+y)^2 dy \right) dx$.

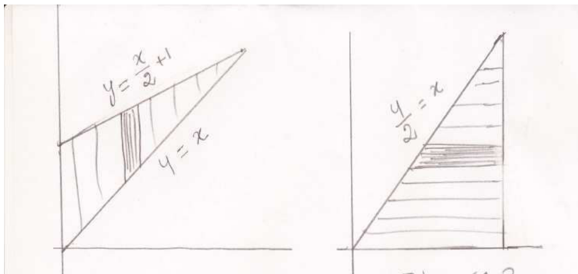


Figure 1 Figure 2

Example 2: Let us evaluate $\int\limits_0^2 \left(\int\limits_{\frac{y}{2}}^1 e^{x^2} dx \right) dy$. Note that we are given two consecutive single integrals. First we have to integrate w.r.to x and then w.r.to y. If we directly integrate then the calculation becomes complicated. So we will use Fubini's theorem and change the order of integration (i.e., $dxdy$ to $dydx$). Note that when we change the order of integration the limits will change.

We will first use Fubini's theorem and convert the consecutive single integrals in to a double integral over a domain D. Note that the integrals are of the form given in the second part of the previous theorem. By the previous theorem (going from right to left) we have $\int_0^2 \left(\int_{y/2}^1 e^{x^2} dx\right) dy = \iint_{D_2} f(x,y)dxdy$ where $D_2 = \{(x,y) : 0 \leq y \leq 2 \ \text{ and } \ \frac{y}{2} \leq x \leq 1\}$ (see Figure 2). Now we will use the first part of the previous theorem and convert this double integral into two consecutive single integrals. By the first part of the previous theorem, $\iint\limits_{D_2} f(x,y)dxdy = \int\limits_0^1 \left(\int\limits_0^{2x} e^{x^2} dy \right) dx = e - 1.$

Lecture 33 : Change of Variable in a Double Integral; Triple Integral

We used Fubini's theorem for calculating the double integrals. We have also noticed that Fubini's theorem can be applied if the domain is in a particular form. In this lecture, we will see that in some cases even if the domain is not in that particular form, using some change of variables, we can transform the original double integral into another double integral over a new region where we can apply Fubini's theorem. This idea is analogous to the method of substitution in single variable:

$$\int_a^b f(x)dx = \int_c^d f[g(t)]g'(t)dt, \tag{1}$$

where $a = g(c)$ and $b = g(d)$.

Change of Variable: Here we deal with the problem of transforming an integral $\iint_S f(x,y)dxdy$ defined over a region S in the xy-plane, into another integral $\iint_T F(x,y)dudv$ defined over a new region T in the uv-plane. Instead of the one function g given in (1), here we have two functions X and Y connecting x, y with u, v as follows: $x = X(u,v)$ and $y = Y(u,v)$. Note that a set of points in the uv-plane is mapped into another set of points in the xy-plane by the maps defined in the above equations.

Basic assumptions: We assume that the mapping from the domain T in the uv-plane to the domain S in the xy-plane is one-one. The functions X and Y are continuous and have continuous partial derivatives $\frac{\partial X}{\partial u}, \frac{\partial X}{\partial v}, \frac{\partial y}{\partial u}$ and $\frac{\partial Y}{\partial v}$. The Jacobian $J(u,v)$ defined below is never zero. The Jacobian is defined as follows:

$$J(u,v) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} \end{vmatrix}$$

The formula: $\iint\limits_S f(x,y)dxdy = \iint\limits_T f[X(u,v), Y(u,v)] \ \mid J(u,v) \mid dudv$

We note that the Jacobian defined in the previous equation and the function $g'(t)$ defined in (1) play similar roles in their respective equations. The proof of the change of variable formula is not easy and so we will not present it here.

Example: Let us find the area of the region S bounded by the hyperbolas $xy = 1$ and $xy = 2$, and the curves $xy^2 = 3$ and $xy^2 = 4$. Note that the area of S is $\iint_S dxdy$. Put $u = xy$ and $v = xy^2$, then $x = \frac{u^2}{v}$ and $y = \frac{v}{u}$. The region T is: $1 \le u \le 2$ and $3 \le v \le 4$. The Jacobian $J(u,v) = \frac{1}{v}$. Therefore by the change of variable formula the area of $S = \iint_S dxdy = \iint_T \frac{1}{v} dudv = \int\limits_3^4 \int\limits_1^2 \frac{1}{v} dudv$.

Special case (Polar coordinate): In this case the variables x and y are changed to r and θ by the following two equations: $x = X(r,\theta) = r\cos\theta$ and $y = Y(r,\theta) = r\sin\theta$. We assume that $r > 0$ and θ lies in $[0, 2\pi)$ or $\theta_0 \le \theta < \theta_0 + 2\pi$ for some θ_0 so that the mapping involved in the change of variable is one-one. The Jacobian is

$$J(u,v) = \begin{vmatrix} \frac{\partial X}{\partial r} & \frac{\partial Y}{\partial r} \\ \frac{\partial X}{\partial \theta} & \frac{\partial Y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

Hence the change of variable formula in this case is : $\iint\limits_S f(x,y)dxdy = \iint\limits_T f(r\cos\theta, r\sin\theta)\ rdrd\theta$.

Example: Let us find the volume of the sphere of radius a. The volume is

$$V = 2\iint\limits_S \sqrt{a^2 - x^2 - y^2}dxdy \quad \text{where} \quad S = \{(x,y) : x^2 + y^2 \le a^2\}.$$

If we use the rectangular coordinates, $V = 4\int\limits_{-a}^{a} \int\limits_{0}^{\sqrt{a^2-x^2}} \sqrt{a^2 - x^2 - y^2}dydx$ which is complicated to calculate. Let use use the polar coordinates. In polar coordinate

$$V = 2\iint\limits_T \sqrt{a^2 - r^2}\, r dr d\theta \quad \text{where} \quad T = [0,a] \times [0, 2\pi].$$

By Fubini's theorem, $V = 2\int\limits_0^a \int\limits_0^{2\pi} \sqrt{a^2-r^2}\, r d\theta dr = 4\pi \int\limits_0^a r\sqrt{a^2-r^2}\, dr = 4\pi \frac{(a^2-r^2)^{\frac{3}{2}}}{-3} \Big|_0^a = \frac{4\pi a^3}{3}.$

Triple integrals: In the previous lecture, we extended the concept of integrals for functions defined on $[a,b] \times [c,d]$. The same can be extended to functions defined on $Q = [a,b] \times [c,d] \times [e,f]$. The definition of integral of such a function is entirely analogous to the definition of double integrals. Every partition P of Q is of the form $P = P_1 \times P_2 \times P_3$ where P_1, P_2 and P_3 are partitions of $[a,b], [b,c]$ and $[e,f]$ respectively. For a given partition P and a bounded function defined on Q we can define $L(P,f), U(P,f)$, lower integral, upper integral and integral of f as we defined in the double integral case. If a function f on Q is integrable then the integral, called triple integral, is denoted by

$$\iiint_Q f(x,y,z) dx dy dz \quad \text{or} \quad \iiint_Q f(x,y,z) dV.$$

As we did in the double integral case, the definition of triple integral can be extended to any bounded region in $\mathbb{R}^3$. One can also prove that every continuous function on Q is integrable.

Remark: In the double integral case, the integral of positive function f is the volume of the region below the surface $z = f(x,y)$. In the triple integral case we do not have any such geometric interpretation, except the fact that $\iiint_D dx dy dz$ is considered to be the volume of the region D. The concept of double integrals can be used in applications, for example, to define the center of mass and moments of inertia of a two dimensional object (see the text book). Similarly, the triple integrals are used in applications which we are not going to see. In this lecture we will see how to evaluate the triple integrals.

There is a Fubini's theorem to evaluate the triple integrals.

Fubini's theorem: Let D be a bounded domain in $\mathbb{R}^3$ described as follows:

$$D = \{(x,y,z) : (x,y) \in R \quad \text{and} \quad f_1(x,y) \le z \le f_2(x,y)\}.$$

That is, D is bounded above by the surface $z = f_1(x,y)$, bounded below by the surface $z = f_2(x,y)$ and on the side by the cylinder generated by a line moving parallel to the z-axis along the boundary of R. The projection of D on the xy-plane is the region R. For example, consider

$$D = \{(x,y,z) : 0 \le x^2 + y^2 \le 2, \; x^2 + y^2 \le z \le 2\}.$$

Here R is a circular region and D is bounded below by the paraboloid $z = x^2 + y^2$ and above by the plane $z = 2$. The following theorem converts a triple integral into iterated integrals of one and two dimensions.

Theorem: If f is continuous on D and f_1, f_2 are continuous on R then

$$\iiint_D f(x,y,z) dV = \iint_R \left(\int_{f_1(x,y)}^{f_2(x,y)} f(x,y,z) dz \right) dA.$$

Example: Let us compute $\iiint_D x dx dy dz$ where D is the region in space bounded by $x = 0, \; y = 0, \; z = 2$ and the surface $z = x^2 + y^2$. Note that $D = \{(x,y,z) : (x,y) \in R, \; x^2 + y^2 \le z \le 2\}$ where $R = \{(x,y) : 0 \le x \le \sqrt{2}, \; 0 \le y \le \sqrt{2 - x^2}\}$. Therefore by the previous theorem,

$$\iiint_D x dx dy dz = \iint_R \left(\int_{x^2+y^2}^2 x dz \right) dA = \int_0^{\sqrt{2}} \int_0^{\sqrt{2-x^2}} \int_{x^2+y^2}^2 x \; dz dy dx = \frac{8\sqrt{2}}{15}.$$

Lecture 34 : Change of Variable in a Triple Integral; Area of a Parametric Surface

The change of variable formula for a double integral can be extended to triple integrals. We will straightaway present the formula.

Formula: $\iiint\limits_S f(x,y,z)dxdydz = \iiint\limits_T f[X(u,v,w),Y(u,v,w),Z(u,v,w)] \mid J(u,v,w) \mid dudvdw$

where the Jacobian determinant $J(u,v,w)$ is defined as follows:

$$J(u,v,w) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} & \frac{\partial Z}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} & \frac{\partial Z}{\partial v} \\ \frac{\partial X}{\partial w} & \frac{\partial Y}{\partial w} & \frac{\partial Z}{\partial w} \end{vmatrix}.$$

The above formula is valid under some assumptions which are similar to the assumptions we had for the two dimensional case.

Special cases : 1. Cylindrical coordinates. In this case the variables x, y and z are changed to r, θ and z by the following three equations:

$$x = X(r,\theta) = r\cos\theta,\ y = Y(r,\theta) = r\sin\theta \quad \text{and} \quad z = z.$$

We assume that $r > 0$ and θ lies in $[0, 2\pi)$ or $\theta_0 \le \theta < \theta_0 + 2\pi$ for some θ_0 as in the double integral case. We have basically replaced x and y by their polar coordinates in the xy plane and left z unchanged. The Jacobian is

$$J(u,v,z) = \begin{vmatrix} \cos\theta & \sin\theta & 0 \\ -r\sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

Therefore the change of variable formula is $\iiint\limits_S f(x,y,z)dxdydz = \iiint\limits_T f(r\cos\theta, r\sin\theta, z)\ rdrd\theta dz$.

Example 1: Let us evaluate $\iiint_D (z^2x^2+z^2y^2)dxdydz$ where D is the region determined by $x^2+y^2 \le 1$, $-1 \le z \le 1$. Note that we can describe D in cylindrical coordinates: $0 \le r \le 1,\ 0 \le \theta \le 2\pi,\ -1 \le z \le 1$. Therefore,

$$\iiint\limits_D (z^2x^2 + z^2y^2)dxdydz = \int_{-1}^{1}\int_{0}^{2\pi}\int_{0}^{1}(z^2r^2)rdrd\theta dz = \int_{-1}^{1}\int_{0}^{2\pi} z^2\frac{r^4}{4}\Big|_{r=0}^{1}\ d\theta dz = \int_{-1}^{1}\frac{2\pi}{4}z^2dz = \frac{\pi}{3}.$$

2 Spherical Coordinates: Suppose (x,y,z) be a point $\mathbb{R}^3$. We will represent this point in terms of spherical coordinates (ρ,θ,ϕ). The coordinates ρ, θ and ϕ are defined below.

Given a point (x,y,z), let $\rho = \sqrt{x^2+y^2+z^2}$ and ϕ is the angle that the position vector $xi+yj+zk$ makes with the (positive side of the) z-axis. The coordinate of z is given by $z = \rho\cos\phi$. To represent x and y in terms of spherical coordinates, represent x and y by polar coordinates in the xy-plane: $x = r\cos\theta$ and $y = r\sin\theta$. Since $r = \rho\sin\phi$, the point (x,y,z) is represented in terms of the spherical coordinates (ρ,θ,ϕ) as follows:

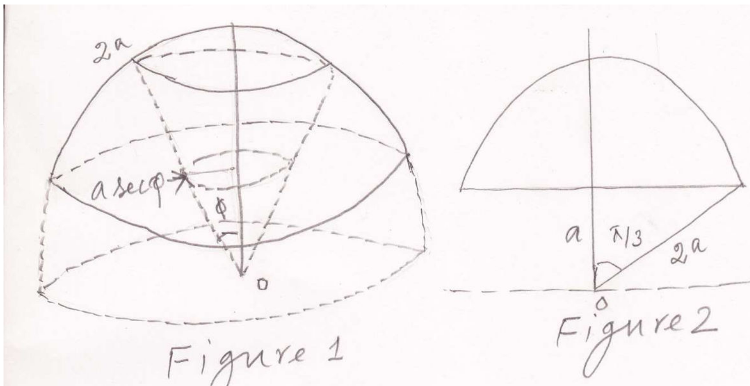
$$x = \rho\sin\phi\cos\theta, \quad y = \rho\sin\phi\sin\theta, \quad z = \rho\cos\phi.$$

We keep $\rho > 0,\ 0 \le \theta < 2\pi$ and $0 \le \phi < \pi$ to get a one-one mapping. The Jacobian determinant is $J(\rho,\theta,\phi) = -\rho^2\sin\phi$. Since $\sin\phi \ge 0$, we have $\mid J(\rho,\theta,\phi) \mid = \rho^2\sin\phi$ and the change of variable formula is

$$\iiint\limits_S f(x,y,z)dxdydz = \iiint\limits_T f(\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\phi)\ \rho^2\sin\phi\ d\rho d\theta d\phi.$$

Example 2: Let $D = \{(x,y,z) : x^2+y^2+z^2 \le 4a^2,\ z \ge a\}$. Let us evaluate $\iiint\limits_D \frac{z}{(x^2+y^2+z^2)^{3/2}}dV$. We will use the spherical coordinates to solve this problem. If we allow ϕ to vary independently, then ϕ varies from 0 to $\frac{\pi}{3}$ (see Figure 2). If we fix ϕ and allow θ to vary from 0 to 2π then we obtain a surface of a cone (see Figure 1). Since only a part of the cone is lying in the given region, for a fixed ϕ and θ, ρ varies from $a\sec\phi$ to $2a$ (see Figure 1). Therefore the integral is

$$\int_0^{\frac{\pi}{3}}\int_0^{2\pi}\int_{a\sec\phi}^{2a} \frac{\cos\phi}{\rho^2} \mid J(\rho,\theta,\phi) \mid d\rho d\theta d\phi = 2\pi\int_0^{\frac{\pi}{3}}(2a\sin\phi\cos\phi - a\sin\phi)d\phi = \frac{\pi a}{2}.$$



Parametric Surfaces: We defined a parametric curve in terms of a continuous vector valued function of one variable. We will see that a continuous vector valued function of two variables is associated with a surface, called parametric surface.

Let T be a region in $\mathbb{R}^2$ and $r(u,v) = X(u,v)i + Y(u,v)j + Z(u,v)k$ be a continuous function on T. The range of r, $\{r(u,v) : (u,v) \in T\}$ is called a parametric surface (with the parameter domain T and the parameters u and v). We assume that the map r is one-one in the interior of T so that the surface does not cross itself. Sometimes the surface defined by $r(u,v)$ is also expressed as

$$x = X(u,v),\ y = Y(u,v),\ z = Z(u,v) \quad \text{where} \quad (u,v) \in T$$

and the above equations are called parametric equations of the surface.

Examples: 1. For a constant $a > 0$, $0 \le \theta \le 2\pi$ and $0 \le \phi \le \pi$ the equations $x = a\sin\phi\cos\theta$, $y = a\sin\phi\sin\theta$, $z = a\cos\phi$ represent a sphere. Here the parameters are θ and ϕ.

2. For a fixed a, $\infty < t < \infty$, $0 \le \theta \le 2\pi$, the equations $x = a\cos\theta$, $y = a\sin\theta$, $z = t$ represent a cylinder. Here the parameters are t and θ.

3. A cone is represented by $r(u,v) = \rho\sin\alpha\cos\theta i + \rho\sin\alpha\sin\theta j + \rho\cos\alpha k$ where $\rho \ge 0$, $0 \le \theta \le 2\pi$ and α is fixed. Here the parameters are ρ and θ.

Area of a Parametric Surface: Let $S = r(u,v)$ be a parametric surface defined on a parameter domain T. Suppose r_u and r_v are continuous on T and $r_u \times r_v$ is never zero on T. Then the area of S, denoted by $a(S)$, is defined by the double integral

$$a(S) = \iint_T \parallel \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \parallel dudv.$$

The formula can be justified as follows. Consider a small rectangle $\triangle A$ in T with the sides on the lines $u = u_0, u = u_0 + \triangle u, v = v_0$ and $v = v_0 + \triangle v$. Consider the corresponding patch in S, that is $r(\triangle A)$. Note that the sides of $\triangle A$ are mapped to the boundary curves of the patch $r(\triangle A)$ by the map r. The vectors $r_u(u_0, v_0)$ and $r_v(u_0, v_0)$ are tangents to the boundary curves of $r(\triangle A)$ meeting at $r(u_0, v_0)$. We now approximate the surface patch $r(\triangle A)$ by the parallelogram whose sites are determined by the vectors $\triangle u r_u$ and $\triangle v r_v$. The area of this parallelogram is $\mid r_u \times r_v \mid \triangle u \triangle v$. This will lead to the Riemann sum corresponding to the double integral $\iint_T \parallel \frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v} \parallel dudv$.

Lecture 35 : Surface Area; Surface Integrals

In the previous lecture we defined the surface area $a(S)$ of the parametric surface S, defined by $r(u, v)$ on T, by the double integral

$$a(S) = \iint_T \| r_u \times r_v \| \, dudv. \tag{1}$$

We will now drive a formula for the area of a surface defined by the graph of a function.

Area of a surface defined by a graph: Suppose a surface S is given by $z = f(x, y)$, $(x, y) \in T$, that is, S is the graph of the function $f(x, y)$. (For example, S is the unit hemisphere defined by $z = \sqrt{1 - x^2 - y^2}$ where (x, y) lies in the circular region $T : x^2 + y^2 \leq 1$.) Then S can be considered as a parametric surface defined by:

$$r(x, y) = xi + yj + f(x, y)k, \quad (x, y) \in T.$$

In this case the surface area becomes

$$a(S) = \iint_T \sqrt{1 + f_x^2 + f_y^2} \, dxdy. \tag{2}$$

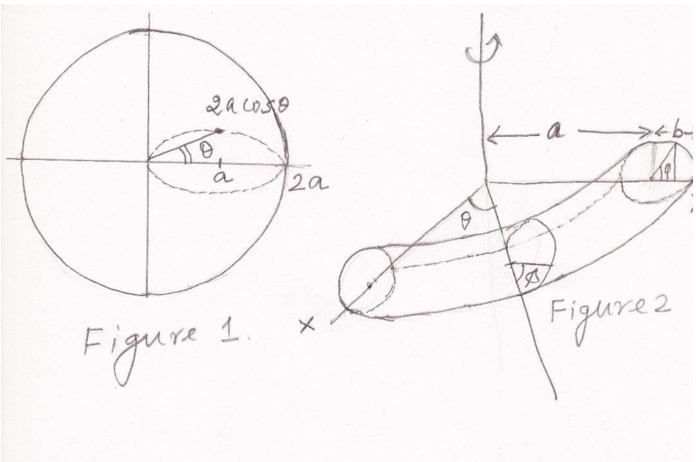
because $\| r_u \times r_v \| = \| -f_x i - f_y j + k \| = \sqrt{1 + f_x^2 + f_y^2}$.

Example 1: Let us find the area of the surface of the portion of the sphere $x^2 + y^2 + z^2 = 4a^2$ that lies inside the cylinder $x^2 + y^2 = 2ax$. Note that the sphere can be considered as a union of two graphs: $z = \pm\sqrt{4a^2 - x^2 - y^2}$. We will use the formula given in (2) to evaluate the surface area. Let $z = f(x, y) = \sqrt{4a^2 - x^2 - y^2}$. Then

$$f_x = \frac{-x}{\sqrt{4a^2 - x^2 - y^2}}, \quad f_y = \frac{-y}{\sqrt{4a^2 - x^2 - y^2}} \quad \text{and} \quad \sqrt{1 + f_x^2 + f_y^2} = \sqrt{\frac{4a^2}{4a^2 - x^2 - y^2}}.$$

Let T be the projection of the surface $z = f(x, y)$ on the xy-plane (see Figure 1). Then, because of the symmetry, the surface area is

$$a(S) = 2 \iint_T \sqrt{\frac{4a^2}{4a^2 - x^2 - y^2}} dxdy = 2 \times 2 \int_0^{\frac{\pi}{2}} \int_0^{2a\cos\theta} \frac{2ardrd\theta}{\sqrt{4a^2 - r^2}}.$$



Remark: Since

$$\| r_u \times r_v \|^2 = \| r_u \|^2 \| r_v \|^2 \sin^2\theta = \| r_u \|^2 \| r_v \|^2 (1 - \cos^2\theta) = \| r_u \|^2 \| r_v \|^2 - (r_u \cdot r_v)^2,$$

the formula given in (1) can be written as

$$a(S) = \iint_T \sqrt{EG - F^2} dudv \tag{3}$$

where $E = r_u \cdot r_u, G = r_v \cdot r_v$ and $F = r_u \cdot r_v$.

Example 2: Let us compute the area of the torus

$$x = (a + b\cos\phi)\cos\theta,\ y = (a + b\cos\phi)\sin\theta,\ z = b\sin\phi$$

where $0 \leq \theta \leq 2\pi,\ 0 \leq \phi \leq 2\pi$, and a and b are constants such that $0 < b < a$. Since the surface is given in the parametric form with the parameters θ and ϕ, we can either use the formula given in (1) or (3) and find the surface area. We do not have to know how the surface looks like. However the surface is given in Figure 2 for understanding. Note that

$$r_\theta = -(a + b\cos\phi)\sin\theta i + (a + b\cos\phi)\cos\theta j + 0k, \quad r_v = -b\sin\phi\cos\theta i - b\sin\phi\sin\theta j + b\cos\phi k.$$

This implies that $E = r_u \cdot r_u = (a + b\cos\phi)^2,\ F = 0, G = b^2$ and hence $\sqrt{EG - F^2} = b(a + b\cos\varphi)$. Therefore, by (3), the surface area is

$$a(S) = \iint_T b(a + b\cos\varphi)d\theta d\phi = \int_0^{2\pi}\int_0^{2\pi} b(a + b\cos\varphi)d\theta d\phi = 4\pi^2 ab.$$

Note that this problem can also be solved using the Pappus theorem : $a(S) = 2\pi\rho L = 2\pi \cdot a \cdot 2\pi b$.

Surface Integrals: We will define the concept of integrals, called surface integrals, to the scalar functions defined on parametric surfaces. Surface integrals are used to define center of mass and moment of inertia of surfaces, and the surface integrals occur in several applications. We will not get in to the applications of the surface integrals in this course. We will define the surface integrals and see how to evaluate them.

Let S be a parametric surface defined by $r(u, v), (u, v) \in T$. Suppose r_u and r_v are continuous. Let $g : S \to \mathbb{R}$ be bounded. The surface integral of g over S, denoted by $\iint_S g d\sigma$, is defined by

$$\iint_S g\ d\sigma = \iint_T g(r(u, v)) \parallel r_u \times r_v \parallel dudv = \iint_T g(r(u, v))\sqrt{EG - F^2}\ dudv \qquad (4)$$

provided the RHS double integral exists. If S is defined by $z = f(x, y)$, then

$$\iint_S g\ d\sigma = \iint_T g[x, y, f(x, y)]\sqrt{1 + f_x^2 + f_y^2}\ dxdy. \qquad (5)$$

where T is the projection of the surface S over the xy-plane.

Example 3: Let S be the hemispherical surface $z = (a^2 - x^2 - y^2)^{1/2}$. Let us evaluate $\iint_S \frac{d\sigma}{[x^2+y^2+(z+a)^2]^{1/2}}$. We first parameterize the surface S as follows:

$$S := r(\theta, \phi) = (a\sin\phi\cos\theta,\quad a\sin\phi\sin\theta,\ a\cos\phi),\quad 0 \leq \theta \leq 2\pi,\quad 0 \leq \phi \leq \pi.$$

Simple calculation shows that $\sqrt{EG - F^2} = a^2\sin\phi$ and $[x^2+y^2+(z+a)^2]^{1/2} = 2a\cos\frac{\phi}{2}$. Therefore, by equation (4), the surface integral is

$$\iint_S \frac{d\sigma}{[x^2+y^2+(z+a)^2]^{1/2}} = \int_0^{2\pi}\int_0^{\pi/2} \frac{a^2\sin\phi}{2a\cos\frac{\phi}{2}}\ d\phi d\theta.$$

Example 4: Let us evaluate the surface integral $\iint_S g\ d\sigma$ where $g(x, y, z) = x+y+z$ and the surface S is described by $z = 2x+3y,\ x \geq 0,\ y \geq 0$ and $x+y \leq 2$. We use the formula given in (5) to evaluate the surface integral. Note that the projection T of the surface is $\{(x, y) : x \geq 0,\ y \geq 0,\ x + y \leq 2\}$. The surface integral is

$$\iint_S g\ d\sigma = \iint_T (x + y + z)\sqrt{1 + f_x^2 + f_y^2}\ dxdy = \int_0^2\int_0^{2-y}(x + y + 2x + 3y)\sqrt{14}\ dxdy.$$

Remark: Under certain general conditions (we deal with surfaces satisfying such conditions) the value of the surface integral is independent of the representation.

Lecture 36: Line Integrals; Green's Theorem

Let $R : [a, b] \to \mathbb{R}^3$ and C be a parametric curve defined by $R(t)$, that is $C(t) = \{R(t) : t \in [a, b]\}$. Suppose $f : C \to \mathbb{R}^3$ is a bounded function. In this lecture we define a concept of integral for the function f. Note that the integrand f is defined on $C \subset \mathbb{R}^3$ and it is a vector valued function. The integral of such a type is called a line integral or a contour integral.

Definition: Suppose R is a differentiable function. The line integral of f along C is denoted by the symbol $\int_C f \cdot dR$ and is defined by

$$\int_C f \cdot dR = \int_a^b f(R(t)) \cdot R'(t)dt$$

provided the RHS integral exists.

The line integrals appear in several physical situations in which the behavior of a vector is studied along a curve such as work done by a force over a curve, flux of the fluid's velocity vector across a curve and so on. We will not deal with such physical situations in this course, however, we will see that the line integrals are useful to calculate certain types of double integrals and areas of plane regions enclosed by parametric curves.

Suppose $f = (f_1, f_2, f_3)$ and $R(t) = (x(t), y(t), z(t))$ then the line integral $\int_C f \cdot dR$ is also written as $\int_C f_1dx + f_2dy + f_3dz$ or $\int_C f_1(x, y, z)dx + f_2(x, y, z)dy + f_3(x, y, z)dz$.

Example 1: Let us compute the line integral $\int_C f \cdot dR$ from $(0, 0, 0)$ to $(1, 2, 4)$ if $f = x^2i + yj + (xz - y)k$

(a) along the line segment joining these two points.

(b) along the curve given parametrically by $x = t^2,\ y = 2t,\ z = 4t^3$.

Solution: (a) Parameterize the line segment as follows: $x = t,\ y = 2t,\ z = 4t$. Then

$$\int_C f \cdot dR = \int_C x^2dx + ydy + (xz - y)dz = \int_0^1 t^2dt + (2t)(2dt) + (4t^2 - 2t)(4dt) = \int_0^1 (17t^2 - 4t)dt = \frac{11}{3}.$$

(b) The parametrization is already given. Repeat the steps given in the solution of (a).

Problem 1: Evaluate $\int\limits_C \frac{-ydx+xdy}{x^2+y^2}$, where $C := \{(x, y) : x^2 + y^2 = r^2\},\ r > 0$.

Solution: Let us consider $C = (rcost, rsint), 0 \le t \le 2\pi$. Then $\int\limits_C \frac{-ydx+xdy}{x^2+y^2} = \int\limits_0^{2\pi} \frac{sin^2t+cos^2t}{sin^2t+cos^2t}dt = 2\pi$

Remark: One can show that a line integral is independent of the parametrization (that preserves the orientation).

The second FTC for line integrals: The second FTC for real functions states that if $f : [a, b] \to \mathbb{R}$ and f' is continuous then $\int_a^b f'(t) = f(b) - f(a)$. This says that the value of the integral (of some function) depends only on the end points and not on the points between them. We will first extend this result to line integrals.

Theorem: Let $S \subset \mathbb{R}^3$, $f : S \to \mathbb{R}$ be differentiable on S and the gradient ∇f be continuous. Let A, B be two points in S. Let $C = \{R(t) : t \in [a, b]\}$ be a curve lying in S and joining the points A and B, that is $R(a) = A$ and $R(b) = B$. Suppose $R't)$ is continuous on $[a, b]$. Then

$$\int_C \nabla f \cdot dR = f(B) - f(A).$$

Proof: Let $g(t) = f(R(t)$. Then

$$\int_C \nabla f \cdot dR = \int_a^b \nabla f(R(t)) \cdot R'(t)dt = \int_a^b g'(t)dt = g(b) - g(a) = f(R(b)) - f(R(a)) = f(B) - f(A).\Box$$

Example 2: Since $\int_C ydx + xdy = \int_C \nabla(xy) \cdot (dx, dy)$, by the previous theorem, the line integral is independent of path joining any two points.

Green's Theorem: The above theorem states that the line integral of a gradient is independent of the path joining two points A and B. Moreover, the line integral of a gradient along a path joining two points A and B is expressed in terms of the values of f at the boundary points A and B. This is analogous to the second FTC of real functions. We will now see a two dimensional analog of the second FTC theorem. It states that a double integral (of certain type of function) over a plane region R can be expressed as a line integral (of some function) along the boundary curve of R. This result is called Green's theorem. To present the formal statement of Green's theorem we need the following definitions.

Let $R : [a, b] \to \mathbb{R}^3$ be continuous.

Simple closed curve: If $R(a) = R(b)$ then the curve described by R is closed. A closed curve such that $R(t_1) \neq R(t_2)$ for every t_1, t_2 in $(a, b]$ is called a simple closed curve.

Piecewise smooth curve: If R' exists and continuous then the curve described by R is called smooth. The curve is called piecewise smooth if the interval $[a, b]$ can be partitioned into a finite number of subintervals and in each of which the curve is smooth.

Theorem: Let C be a piecewise smooth simple closed curve in the xy-plane and let D denote the closed region enclosed by C. Suppose $M, N, \frac{\partial N}{\partial x}$ and $\frac{\partial M}{\partial y}$ are real valued continuous functions in an open set containing D. Then

$$\iint_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dxdy = \oint_C (Mi + Nj) \cdot dR = \oint_C Mdx + Ndy$$

where the line integral is taken around C in the counterclockwise direction.

Remark: In the above theorem we have made some casual statements such as "closed region enclosed by C" and "counterclockwise direction". These are intuitively evident, however, formal definitions and some explanations are required which we are not going to provide. We will also make a few such statements in the next two or three lectures. So we have to be aware that our treatment is not completely rigorous. Proof of the previous theorem for certain special regions is given in the text book and the proof in the general form is not easy.

An application. *Area expressed as a line integral:* Let C be a simple (piecewise smooth) closed curve and D be the region enclosed by C. Let $N(x, y) = \frac{x}{2}$ and $M(x, y) = -\frac{y}{2}$, then by Green's theorem the area of D is

$$a(D) = \iint_D dxdy = \iint_D (N_x - M_y)dxdy = \int_a^b Mdx + Ndy = \frac{1}{2} \int_C -ydx + xdy.$$

Examples: 1. Let us show that the value of $\int_C xy^2dx + (x^2y + 2x)dy$ around any square depends only on the size of the square C and not on its location in the plane. Let R be a square enclosed by the boundary C. By Green's theorem

$$\int_C xy^2dx + (x^2y + 2x)dy = \iint_R 2dxdy = 2 \text{ Area(R)}.$$

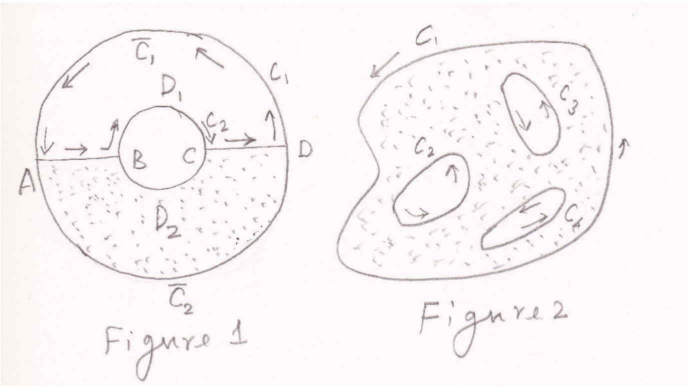
2. We will use the formula given above to find the area bounded by the ellipse $C : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Parametrize C by $(a \cos t, b \sin t),\ 0 \leq t \leq 2\pi$. Then the area is

$$\frac{1}{2} \int_C -ydx + xdy = \frac{1}{2} \int_0^{2\pi} -(b \sin t)(-a \sin t)dt + (a \cos t)(b \cos t)dt = \frac{1}{2} \int_0^{2\pi} abdt = ab\pi.$$

Lecture 37: Green's Theorem (contd.); Curl; Divergence

We stated Green's theorem for a region enclosed by a simple closed curve. We will see that Green's theorem can be generalized to apply to annular regions.

Suppose C_1 and C_2 are two circles as given in Figure 1. Consider the annular region (the region between the two circles) D. Introduce the crosscuts AB and CD as shown in Figure 1. Consider the simple closed curve $\overline{C_1}$ consisting of the upper half of C_2, the upper half of C_1, and the segments AB and CD as shown in Figure 1. Similarly, consider the simple closed curve $\overline{C_2}$ consisting of the lower half of C_2, the lower half of C_1, and the segments AB and CD. Let D_1 and D_2 be the regions enclosed by $\overline{C_1}$ and $\overline{C_2}$.



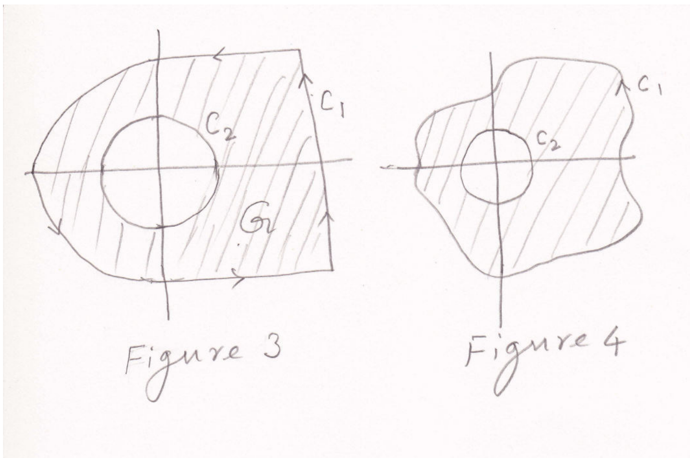
Suppose we are given two continuously differentiable scalar valued functions M and N on an open set containing the annular region D. Let us now apply Green's theorem to each of the regions D_1 and D_2 and add the two identities obtained from Green's theorem. Since the line integrals along the crosscuts cancel, we obtain

$$\iint_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dxdy = \oint_{C_1}(Mdx + Ndy) - \oint_{C_2}(Mdx + Ndy). \tag{1}$$

In the above equation, the line integrals are taken around the curves in the counterclockwise directions. Note that when we apply Green's theorem on D_1, the line integral on the part of $\overline{C_2}$ is taken along the clockwise direction. So a minus sign appears in the above equation.

We note that using the idea given above we can generalize Green's theorem to apply to regions enclosed by two or more simple closed curves similar to the one given in Figure 2.

Example 1: Let G be the region outside the unit circle which is bounded on left by the parabola $y^2 = 2(x+2)$ and on the right by the line $x = 2$. Use Green's theorem to evaluate $\int_{C_1} \frac{xdy - ydx}{x^2+y^2}$ where C_1 is the outer boundary of G oriented counterclockwise.



Solution: Let C_2 be the unit circle (see Figure 3). If we take $M = -\dfrac{y}{x^2+y^2}$ and $N = \dfrac{x}{x^2+y^2}$, then a simple calculation shows that $N_x - M_y = 0$. Therefore $\iint_G (N_x - M_y)dx\,dy = 0$. By applying Green's theorem on G (as we did above to obtain (1)), we get

$$\oint_{C_1}(Mdx + Ndy) = \oint_{C_2}(Mdx + Ndy)$$

where the line integrals around both the curves are taken in the counterclockwise directions. We have already seen in Problem 1 of the previous lecture that $\oint_{C_2}(Mdx + Ndy) = 2\pi$. □

Remark: If we take C_2 be any circle centered at $(0,0)$ and C_1 be any (piecewise smooth) simple closed curve such that C_2 lies in the interior of C_1 as shown in Figure 4, by repeating the argument given in the above solution, we can show that $\oint_{C_1}(Mdx + Ndy) = 2\pi$ where M and N are given in the previous example.

Problem: Evaluate $\int_C \frac{xdy - ydx}{x^2+y^2}$ along any simple closed curve C in the xy plane not passing through the origin. Distinguish the cases where the region D enclosed by C: (a) includes the origin (b) does not include the origin.

Solution: (a) First note that if we take $M = -\dfrac{y}{x^2+y^2}$ and $N = \dfrac{x}{x^2+y^2}$, then the functions are not defined in the region D, hence one cannot apply Green's theorem. Choose a circle C_r of radius r centered at $(0,0)$ and C_r lies in the interior of C. Now one can apply Green's theorem on the region between these two curves. By the above remark, the value of the line integral is 2π.

(b) In this region we can apply Green's theorem. Therefore $\int\limits_C Mdx + Ndy = \iint\limits_D (N_x - M_y)dx\,dy = 0$.

Curl and divergence: In the previous two lectures we discussed Green's theorem which expresses a double integral (of certain type of function) over a plane region D as a line integral over the boundary of D. We have also noted that this is a two dimensional analog of the second FTC. In the next two lectures we will see two generalizations of Green's theorem involving surface integrals and triple integrals. These results are known as Stokes theorem and divergence theorem respectively. They are also, essentially, analogs of the second FTC.

We first rewrite Green's theorem into two different forms involving the concepts curl and divergence and then generalize these forms to surface integrals and triple integrals. Let us define the concepts curl and divergence.

Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ and $F(x,y,z) = P(x,y,z)i + Q(x,y,z)j + R(x,y,z)k$. Such functions are called vector field.

Curl: The curl of F is another vector field denoted by $curlF$ and defined by

$$curlF = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) i + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) j + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) k.$$

We rewrite the curl as follows: $curlF = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = \nabla \times f$. These expressions can be easily remembered. However, while expanding the determinant it is understood that $\frac{\partial}{\partial x}$ times Q is to be interpreted as $\frac{\partial Q}{\partial x}$ and the symbol ∇ has to be treated as if it is vector $\nabla = \frac{\partial}{\partial x}i + \frac{\partial}{\partial y}j + \frac{\partial}{\partial z}k$.

Divergence: The divergence of F is a scalar valued function denoted by $divF$ and is defined by

$$divF = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z}.$$

We can rewrite the div F as follows : $divF = \nabla \cdot F$. Note that we interpret $\frac{\partial}{\partial x}$ times Q as $\frac{\partial Q}{\partial x}$.

Lecture 38: Stokes' Theorem

As mentioned in the previous lecture Stokes' theorem is an extension of Green's theorem to surfaces. Green's theorem which relates a double integral to a line integral states that

$$\iint\limits_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y} \right) dxdy = \oint_C Mdx + Ndy$$

where D is a plane region enclosed by a simple closed curve C. Stokes' theorem relates a surface integral to a line integral. We first rewrite Green's theorem in a different form as mentioned in the previous lecture. Consider the vector field $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(x, y) = M(x, y)i + N(x, y)j$. Then $curlF = \nabla \times F = \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y} \right) k$. Therefore Green's theorem is stated as follows:

$$\iint_D (curlF) \cdot k \ dxdy = \oint_C F \cdot dR. \tag{1}$$

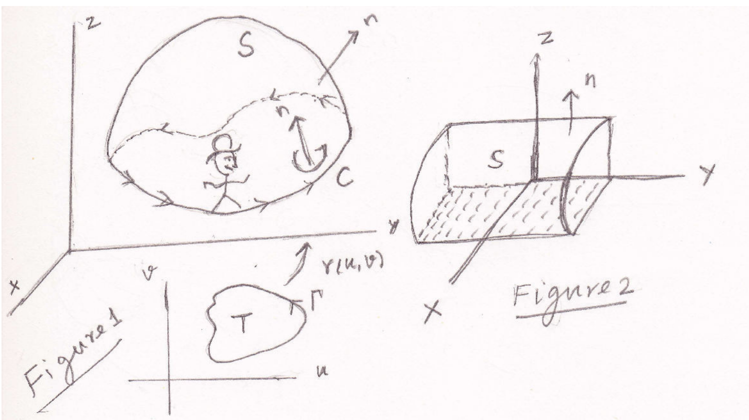
We will extend this form of Green's theorem to a vector field F defined on a surface S having the boundary curve C. We need the following definitions.

Smooth Surface: Let $S = r(u, v)$ be a parametric surface defined on a parameter domain T. We say that S is smooth if r_u and r_v are continuous on T and $r_u \times r_v$ is never zero on T. A level surface S defined by $f(x, y, z) = c$ is said to be smooth if ∇f is continuous and never zero on S

We have already seen that the vectors $r_u \times r_v$ and ∇f are normals to the parametric surface and the level surface respectively.

Orientable Surface: A smooth surface is said to be orientable if there exists a continuous unit normal vector function defined at each point of the surface.

Basically a surface is oriented by orienting its normals in a continuous manner. In practice, we consider an orientable surface as a smooth surface with two sides. For example spheres, planes and paraboloids are orientable surfaces. The Möbious strip is not an orientable surface and is not one sided.



We will be dealing with only orientable surfaces. Let S be an orientable surface with the boundary curve C (see Figure 1). Since S has two sides, consider a side of the surface, that is, consider a normal $\mathbf{n}$ to the surface S. With respect to this normal (that is corresponding to a side of the surface), we define an orientation on C. We need this orientation to evaluate the line integral involved in Stokes' theorem.

Orientation on the boundary w.r.to a normal: Orientation is formally defined as follows. Suppose, for example, $S := r(u, v)$ is a parametric (orientable) surface defined by a one-one map r on the parameter domain T. Consider the unit normal $\mathbf{n} = \frac{r_u \times r_v}{\|r_u \times r_v\|}$ of the surface S. Let Γ be the boundary of T and $C = r(\Gamma)$ (see Figure 1). If we assume that Γ is oriented in the counterclockwise direction then we get an orientation for C inherited from Γ through the mapping r. This orientation for C

is considered to be the orientation w.r.to **n**.

In practice, we get the orientation of the boundary curve (w.r.to a given **n**) using the right-hand rule: If we wrap around the surface with the right hand by pointing the thumb in the direction of the normal then roughly the fingers of the right hand are pointing towards the orientation of the curve. One can also use the following method. Consider a person walking on C by keeping his or her head towards the direction of the normal and the surface to the left. The orientation of the curve is the direction in which the person is walking on C (see Figure 1). For example, if S is a plane region and $\mathbf{n} = k$ then the orientation of C w.r.to **n** is the counterclockwise direction. Let us state Stokes' theorem.

Theorem: Let S be a (piecewise) smooth orientable surface and a piecewise simple closed curve C be its boundary. Let $F(x, y, z) = P(x, y, z)i + Q(x, y, z)j + R(x, y, z)k$ be a vector field such that P, Q and R are continuous and have continuous first partial derivatives in an open set containing S. If **n** is a unit normal to S, then

$$\iint_S (curl F) \cdot n \; d\sigma = \oint_C F \cdot dR \tag{2}$$

where the line integral is taken around C in the direction of the orientation of C w.r.to **n**.

Remarks: 1. The value of the surface integral in (2) depends only on the boundary C. This means that the shape of the surface is irrelevant. So Stokes theorem is an analog of the 2nd FTC.

2. If S is a plane region, then the identity given in (2) reduces to the identity given in (1). Therefore Stokes' theorem is consider to be a direct extension of Green's theorem.

3. For a closed oriented surface such as sphere or donut, there is no boundary and in this case $\iint_S (curl F) \cdot n \; d\sigma = 0$. For example for a sphere, this can be seen by cutting the sphere into two hemispheres. If we apply Stokes' theorem to each and add the resulted identities, the two boundary integrals cancel and we get what we claimed.

4. Stokes' theorem can also be extended to a smooth surface which has more than one simple closed curve forming the boundary of the surface.

Problem: Let S be the part of the cylinder $z = 1 - x^2, \; 0 \le x \le 1, \; -2 \le y \le 2$. Let C be the boundary curve of the surface S. Let $F(x, y, z) = yi + yj + zk$. Find the unit outer normal to S, *curl F* and evaluate $\oint_C F \cdot dR$ where C is oriented counterclockwise as viewed from above the surface.

Solution: The surface is given by $z = 1 - x^2$ (see Figure 2). This surface can be considered as a graph of the function $f(x, y) = 1 - x^2$ or a parametric surface $r(x, y) = (x, y, f(x, y))$. An unit normal is $\frac{r_x \times r_y}{\|r_x \times r_y\|} = \frac{-f_x - f_y + k}{\sqrt{1+f_x^2+f_y^2}} = \frac{2xi+k}{\sqrt{1+4x^2}}$. This has to be the outer normal, because if we calculate at $(0, 1)$, it coincides with the outer normal k. Simple calculation shows that $curl F = -k$. We will use Stokes' theorem to solve this problem. Note that $curl f \cdot \mathbf{n} = \frac{-1}{\sqrt{1+4x^2}}$. By Stokes' theorem

$$\oint_C F \cdot dR = \iint_S \frac{-1}{\sqrt{1+4x^2}} \; d\sigma = \iint_R \frac{-1}{\sqrt{1+4x^2}} \sqrt{1 + f_x^2 + f_y^2} dxdy = \iint_R -1 dxdy = \int_0^1 \int_{-2}^2 -1 dydx. \; \Box$$

Remarks: 1. The above problem can be done directly by calculating the line integral. In that case we have to parametrize the boundary which consists of four smooth curves. One has to be careful with the direction in each piece. The calculation also becomes lengthy.

2. We can also consider the surface given in the previous problem as a level surface defined by $f(x, y, z) = z - 1 + x^2 = 0$. In this case an unit normal is $\mathbf{n} = \frac{\nabla f}{\|\nabla f\|} = \frac{2xi+k}{\sqrt{1+4x^2}}$.

Lecture 39: The Divergence Theorem

In the last few lectures we have been studying some results which relate an integral over a domain to another integral over the boundary of that domain. In this lecture we will study a result, called divergence theorem, which relates a triple integral to a surface integral where the surface is the boundary of the solid in which the triple integral is defined.

Divergence theorem is a direct extension of Green's theorem to solids in $\mathbb{R}^3$. We will now rewrite Green's theorem to a form which will be generalized to solids.

Let D be a plane region enclosed by a simple smooth closed curve C. Suppose $F(x,y) = M(x,y)i + N(x,y)j$ is such that M and N satisfy the conditions given in Green's theorem. If the curve C is defined by $R(t) = x(t)i + y(t)j$ then the vector $\mathbf{n} = \frac{dy}{ds}i - \frac{dx}{ds}j$ is a unit normal to the curve C because the vector $T = \frac{dx}{ds}i + \frac{dy}{ds}j$ is a unit tangent to the curve C. By Green's theorem

$$\oint_C (F \cdot \mathbf{n})ds = \oint_C Mdy - Ndx = \iint_D \left(\frac{\partial M}{\partial x} - (-\frac{\partial N}{\partial y})\right) dxdy = \iint_D \left(\frac{\partial M}{\partial x} + \frac{\partial N}{\partial y}\right) dxdy.$$

Since $divF = \frac{\partial M}{\partial x} + \frac{\partial N}{\partial y}$, Green's theorem takes the following form:

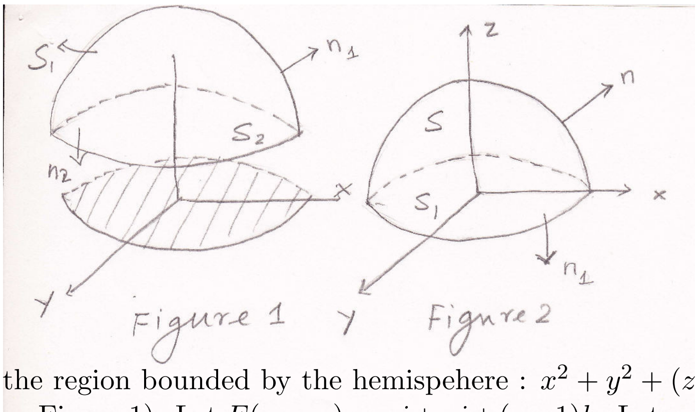
$$\iint_D divF dxdy = \oint_C (F \cdot \mathbf{n})ds.$$

We will now generalize this form of Green's theorem to a vector field F defined on a solid.

Theorem: Let D be a solid in $\mathbb{R}^3$ bounded by piecewise smooth (orientable) surface S. Let $F(x,y,z) = P(x,y,z)i + Q(x,y,z)j + R(x,y,z)k$ be a vector field such that P, Q and R are continuous and have continuous first partial derivatives in an open set containing D. Suppose $\mathbf{n}$ is the unit outward normal to the surface S. Then

$$\iiint_D divF \; dV \;\; = \;\; \iint_S F \cdot \mathbf{n} \; d\sigma.$$

Remark: The divergence theorem can be extended to a solid that can be partitioned into a finite number of solids of the type given in the theorem. For example, the theorem can be applied to a solid D between two concentric spheres as follows. Split D by a plane and apply the theorem to each piece and add the resulting identities as we did in Green's theorem.



Example: Let D be the region bounded by the hemispehere : $x^2 + y^2 + (z-1)^2 = 9, \; 1 \le z \le 4$ and the plane $z = 1$ (see Figure 1). Let $F(x,y,z) = xi + yj + (z-1)k$. Let us evaluate the integrals given in the divergence theorem.

Triple integral: Note that $divF = 3$. Therefore,

$$\iiint_D divF \; dV = \iiint_D 3 \; dV = 3 \cdot \frac{2}{3}\pi 3^3 = 54\pi.$$

Surface integral: The solid D is bounded by a surface S consisting of two smooth surfaces S_1 and S_2 (see Figure 1). Therefore

$$\iint_S F \cdot \mathbf{n}\ d\sigma = \iint_{S_1} F \cdot \mathbf{n}_1\ d\sigma + \iint_{S_2} F \cdot \mathbf{n}_2\ d\sigma.$$

Surface integral over the hemisphere S_1: The surface S_1 is given by:

$$g(x,y,z) = x^2 + y^2 + (z-1)^2 - 9 = 0.$$

An unit normal is

$$\mathbf{n}_1 = \frac{\nabla g}{\| \nabla f \|} = \frac{xi + yj + (z - 1k}{\sqrt{x^2 + y^2 + (z-1)^2}} = \frac{x}{3}i + \frac{y}{3}j + \frac{(z-1)}{3}k.$$

This is expected because the position vector is a normal to the sphere. It is clear that the normal obtained is the outward normal. This implies that over S_1,

$$F \cdot \mathbf{n}_1 = (x, y, z-1) \cdot (\frac{x}{3}, \frac{y}{3}, \frac{(z-1)}{3}) = 3.$$

Therefore

$$\iint\limits_{S_1} F \cdot n_1\ d\sigma = 3 \iint\limits_{S_1} d\sigma = 3 \cdot (surface\ area) = 3 \cdot 18\pi = 54\pi.$$

Surface integral over the plane region S_2: Here the outward normal $\mathbf{n}_2 = -k$. Therefore $F \cdot \mathbf{n}_2 = -z + 1$. Since on S_2, $z = 1$

$$\iint\limits_{S_2} F \cdot \mathbf{n}_2\ d\sigma = 0.$$

Hence $\iint_S F \cdot \mathbf{n}\ d\sigma = 54\pi$.

Problem: Use the divergence theorem to evaluate the surface integral $\iint_S F{\cdot}n\ d\sigma$ where $F(x,y,z) = (x+y, z^2, x^2)$ and S is the surface of the hemisphere $x^2 + y^2 + z^2 = 1$ with $z > 0$ and $\mathbf{n}$ is the outward normal to S.

Solution: First note that the surface is not closed. If we apply the divergence theorem to the solid $D := x^2 + y^2 + z^2 \leq 1,\ z > 0$, we get

$$\iiint\limits_D divF\ dV \ = \iint\limits_S F \cdot \mathbf{n}\ d\sigma + \iint\limits_{S_1} F \cdot \mathbf{n}_1\ d\sigma$$

where $S_1 := x^2 + y^2 < 1,\ z = 0$ the base of the hemisphere (see Figure 2) and $\mathbf{n}_1$ is the outward normal to S_1 which is $-k$. Since $divF = 1$, the volume integral in the above equation is the volume of the hemisphere, $\frac{2\pi}{3}$. Therefore

$$\iint\limits_S F \cdot \mathbf{n}\ d\sigma = \frac{2\pi}{3} - \iint\limits_{S_1} F \cdot -k\ d\sigma = \frac{2\pi}{3} + \iint\limits_{S_1} x^2\ d\sigma$$

which is relatively easier to evaluate. To evaluate the surface integral over S_1, consider $S_1 = (\cos\theta, r\sin\theta),\ 0 \leq r < 1,\ 0 \leq \theta \leq 2\pi$. Then

$$\iint\limits_{S_1} x^2\ d\sigma = \int_0^1 \int_0^{2\pi} r^2 \cos^2\theta r d\theta dr = \int_0^1 r^3 \pi dr = \frac{\pi}{4}.$$

Therefore the requied integral is

$$\iint\limits_S F \cdot \mathbf{n}\ d\sigma = \frac{2\pi}{3} + \frac{\pi}{4} = \frac{11\pi}{12}.$$