MA109 Short Notes

Aditya Byju

Course Professor: Prof. Ravi Raghunathan
Ref: Prof's slides
Why are pirates the best at calculus?
because a true pirate never forgets the c

Calculus - 1

September 2021

Contents

Introduction

- A sequence in a set X is a function $a : \mathbb{N} \to X$, that is, a function from the natural numbers to X
- Sequence of partial sums is given by $\{s_n\}_{n=1}^{\infty}$, where s_n is:

$$s_n = \sum_{k=1}^{n} a_k$$

- **Def:** A sequence is said to be a monotonically increasing sequence if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$
- **Def:** A sequence is said to be a monotonically decreasing sequence if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}$
- A monotonic sequence is one that is either monotonically increasing or monotonically decreasing
- A sequence is called eventually monotonically decreasing when it becomes monotonically increasing or decreasing after some stage

Limits

- **Def:** A sequence a_n tends to a limit l/converges to a limit l, if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that:

  $$|a_n - l| < \epsilon$$

  whenever $n > N$. This is what we mean when we write:

  $$\lim_{n \to \infty} a_n = l$$

- If a sequence has a limit it is said to be convergent. A sequence that does not converge is said to diverge, or to be divergent.
- **The Sandwich theorems:**
  - Theorem 1: If a_n, b_n and c_n are convergent sequences such that $a_n \leq b_n \leq c_n$ for all n, then:

    $$\lim_{n \to \infty} a_n \leq \lim_{n \to \infty} b_n \leq \lim_{n \to \infty} c_n$$

  - Theorem 2: Suppose $\lim_{n \to \infty} a_n = \lim_{n \to \infty} c_n$. If b_n is a sequence satisfying $a_n \leq b_n \leq c_n$ for all n, then b_n converges and:

    $$\lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n = \lim_{n \to \infty} c_n$$

    Note that in the second theorem we do not assume that b_n converges, thus we get the convergence of b_n for free
- **Def:** A sequence a_n is said to be bounded if there is a real number $M > 0$ such that $|a_n| \leq M$ for every $n \in N$. A sequence that is not bounded is called unbounded.
- Bounded sequences don't necessarily converge (for e.g. $a_n = (-1)^n$)
- **Lemma:** Every convergent sequence is bounded
- **Def:** A sequence a_n is said to be bounded above (resp. bounded below) if $a_n < M$ (resp. $a_n > M$) for some $M \in \mathbb{R}$. A sequence that is bounded both above and below is obviously bounded.

- **Theorem:** A monotonically increasing (resp. decreasing) sequence which is bounded above (resp. below) converges

- The limit of a monotonically increasing sequence a_n bounded above is the supremum or least upper bound (lub) of the sequence

- The limit of a monotonically decreasing sequence a_n bounded below is the infimum or greatest lower bound (glb) of the sequence

- A sequence bounded above may not have a maximum but will always have a supremum

- If we change finitely many terms of a sequence it does not affect the convergence and boundedness properties of a sequence. If it is convergent, the limit will not change. If it is bounded, it will remain bounded though the supremum may change.

- **Def:** a sequence a_n in $\mathbb{R}$ is said to be a Cauchy sequence if for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that:

  $$|a_n - a_m| < \epsilon$$

  for all $m,\ n > N$

- **Theorem:** every Cauchy sequence in $\mathbb{R}$ converges (to a real number)

- **Theorem:** every convergent sequence (in any set X) is Cauchy

- A set X in which every Cauchy sequence converges (to a limit in X) is called a complete set. The real numbers are complete.

- Two sequences a_n and b_n will be related to each other (and we write $a_n \sim b_n$) if:

  $$\lim_{n \to \infty} |a_n - b_n| = 0$$

  This is an equivalence relation and it is a fact that it partitions the set S into disjoint classes. The set of disjoint classes is denoted by $S/\sim$. If two sequences converge to the same limit, they are necessarily in the same class. A real number is an equivalence class in $S/\sim$. So a real number should be thought of as the collection of all rational sequences which converge to it.

- **Achilles and the tortoise (Zeno's paradox):** in a race, the quickest runner can never overtake the slowest, since the pursuer must first reach the point whence the pursued started, so that the slower must always hold a lead

- **Def:** A function $f : (a, b) \to \mathbb{R}$ is said to tend to (or converge to) a limit l at a point $x_0 \in [a, b]$ if for all $\epsilon > 0$ there exists $\delta > 0$ such that:

  $$|f(x) - l| < \epsilon$$

  for all $x \in (a, b)$ such that $0 < |x - x_0| < \delta$. In this case, we write:

  $$\lim_{x \to x_0} f(x) = l$$

- The limit of a function may exist even if the function is not defined at that point

- **The Sandwich theorems (for functions):**

  – **Theorem 1:** As $x \to x_0$, if $f(x) \to l_1$, $g(x) \to l_2$ and $h(x) \to l_3$ for functions $f,\ g,\ h$ on some interval (a, b) such that $f(x) \leq g(x) \leq h(x)$ for all $x \in (a, b)$, then:

  $$l_1 \leq l_2 \leq l_3$$

- **Theorem 2:** Suppose $\lim_{x \to x_0} f(x) = \lim_{x \to x_0} h(x) = l$ and If $g(x)$ is a function satisfying $f(x) \leq g(x) \leq h(x)$ for all $x \in (a,b)$, then $g(x)$ converges to a limit as $x \to x_0$ and:

$$\lim_{x \to x_0} g(x) = l$$

Note that in the second theorem we do not assume that $g(x)$ converges, thus we get the convergence of $g(x)$ for free

- **Lemma:** let $f : (a,b) \to \mathbb{R}$ be a function such that $lim_{x \to c} f(x)$ exists for some $c \in [a,b]$. If $c \in (a,b)$, there exists an (open) interval $I = (c-\eta, c+\eta) \subset (a,b)$ such that $f(x)$ is bounded on I. If $c = a$, then there is an open interval $I_1 = (a, a+\eta)$ such that $f(x)$ is bounded on I_1. Similarly if $c = b$, there exists an open interval $I_2 = (b-\eta, b)$ such that $f(x)$ is bounded on I_2.

- **Def:** We say that $f : \mathbb{R} \to \mathbb{R}$ tends to a limit l as $x \to \infty$ (resp. $x \to -\infty$) if for all $\epsilon > 0$ there exists $X \in \mathbb{R}$ such that:

$$|f(x) - l| < \epsilon$$

whenever $x > X$ (resp. $x < X$), and we write:

$$\lim_{x \to \infty} f(x) = l \ \ or \ \lim_{x \to -\infty} f(x) = l$$

Continuity

- **Def:** if $f : [a,b] \to \mathbb{R}$ is a function and $c \in [a,b]$, then f is said to be continuous at the point c if and only if $\lim_{x \to c} f(x) = f(c)$
- A function f on (a,b) (resp. $[a,b]$) is said to be continuous if and only if it is continuous at every point c in (a,b) (resp. $[a,b]$). If f is not continuous at a point c we say that it is discontinuous at c, or that c is a point of discontinuity for f.
- Rational functions are functions of the form $R(x) = P(x)/Q(x)$ where $P(x)$ and $Q(x)$ are polynomials
- **Theorem:** let $f : (a,b) \to (c,d)$ and $g : (c,d) \to (e,f)$ be functions such that f is continuous at x_0 in (a,b) and g is continuous at $f(x_0) = y_0$ in (c,d). Then the function $g(f(x))$ (also written as $g \circ f(x)$ sometimes) is continuous at x_0. So the composition of continuous functions is continuous.
- **The intermediate value theorem:** Suppose $f : [a,b] \to \mathbb{R}$ is a continuous function. For every u between $f(a)$ and $f(b)$ there exists $c \in [a,b]$ there such that $f(c) = u$. Functions which have this property are said to have the Intermediate Value Property (IVP).
- **Theorem:** every polynomial of odd degree has at least one real root
- **The extreme value theorem:** a continuous function on a closed bounded interval $[a,b]$ is bounded and attains its infimum and supremum, that is, there are points x_1 and x_2 in $[a,b]$ such that $f(x_1) = m$ and $f(x_2) = M$, where m and M denote the infimum and supremum respectively
- **Theorem:** a function $f(x)$ is continuous at a point a if and only if for every sequence $x_n \to a$, $\lim_{x_n \to a} f(x_n) = f(a)$. A function that satisfies the above property is said to be sequentially continuous.
- **Theorem:** a function $f : (a,b) \to \mathbb{R}$ is continuous at c if and only if it is sequentially continuous at c
- Functions that satisfy the following property for some α (not necessarily greater than 1) are said to be Lipschitz continuous with exponent α:

$$\left| \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \right| \leq C \lim_{h \to 0} |h|^{\alpha - 1} = 0$$

- **Def:** the function f is said to attain a maximum (resp. minimum) at a point $x_0 \in X$ if $f(x) \leq f(x_0)$ (resp. $f(x) \geq f(x_0)$) for all $x \in X$
- **Def:** let $f : X \to \mathbb{R}$ be a function and x_0 be in X. Suppose there is a sub-interval $x_0 \in (c, d) \subset X$ such that $f(x_0) \geq f(x)$ (resp. $f(x_0) \leq f(x)$) for all $x \in (c, d)$, then f is said to have a local maximum (resp. local minimum) at x_0

Differentiation

- **Fermat's theorem:** : if $f : X \to \mathbb{R}$ is differentiable and has a local minimum or maximum at a point $x_0 \in X$, $f'(x_0) = 0$
- **Rolle's theorem:** Suppose $f : [a, b] \to \mathbb{R}$ is a continuous function which is differentiable in (a, b) and $f(a) = f(b)$. Then there is a point x_0 in (a, b) such that $f'(x_0) = 0$
- If $P(x)$ is a polynomial of degree n with n real roots, then all the roots of $P'(x)$ are also real
- **The mean value theorem:** Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function and that f is differentiable in (a, b). Then there is a point x_0 in (a, b) such that:

$$\frac{f(b) - f(a)}{b - a} = f'(x_0)$$

- Rolle's theorem is a special case of the Mean Value Theorem (MVT)
- **Theorem:** if f satisfies the hypotheses of the MVT, and further $f'(x_0) = 0$ for every $x \in (a, b)$, then f is a constant function
- **Darboux's theorem:** Let $f : (a, b) \to \mathbb{R}$ be a differentiable function. If $c,\ d,\ c < d$ are points in (a, b), then for every u between $f'(c)$ and $f'(d)$, there exists an x in $[c, d]$ such that $f'(x) = u$
- A point x_0 in (a, b) such that $f'(x_0) = 0$ is often called a stationary point
- **Second derivative test:** assume that $f : [a, b] \to \mathbb{R}$ is a continuous function and that f is differentiable on (a, b). Also assume that $f'(x)$ is differentiable at x_0, that is, that the second derivative $f''(x_0)$ exists. Then:
  - If $f''(x_0) > 0$, the function has a local minimum at x_0
  - If $f''(x_0) < 0$, the function has a local maximum at x_0
  - If $f''(x_0) = 0$, no conclusion can be drawn
- **Def:** a point of inflection x_0 for a function f is a point where the function changes its behavior from concave to convex (or vice-versa). At such a point $f''(x_0) = 0$, but this is only a necessary, and not a sufficient condition.
- **Def:** let I denote an interval (open or closed or half-open). A function $f : I \to \mathbb{R}$ is said to be concave (or sometimes concave downwards) if:

$$f(tx_1 + (1 - t)x_2) \geq tf(x_1) + (1 - t)f(x_2)$$

  for all x_1 and x_2 in I and $t \in [0, 1]$.
  Similarly, a function is said to be convex (or concave upwards) if:

$$f(tx_1 + (1 - t)x_2) \leq tf(x_1) + (1 - t)f(x_2)$$

  By replacing the $\geq$ and $\leq$ signs above by strict inequalities we can define strictly concave and strictly convex functions.

- Every convex function is Lipschitz continuous with $\alpha = 1$
- A convex function is differentiable at all but at most countably many points
- A twice differentiable function on an interval will be convex if its second derivative is everywhere non-negative. If the second derivative is positive, the function will be strictly convex.
- The space $\mathcal{C}^k(I)$, will denote the space of k times continuously differentiable functions on an (open) interval I, for some fixed $k \in \mathbb{N}$, that is, the space of functions for which k derivatives exist and such that the k^{th} derivative is a continuous function. The space $\mathcal{C}^\infty(I)$ will consist of functions that lie in $\mathcal{C}^k(I)$ for every $k \in \mathbb{N}$. Such functions are called smooth or infinitely differentiable functions.
- **The Taylor polynomials:** given a function $f(x)$ which is n times differentiable at some point x_0 in an interval I, we can associate to it a family of polynomials $P_0(x),\ P_1(x), \ldots,\ P_n(x)$ called the Taylor polynomials of degrees $0,\ 1, \ldots\ n$ at x_0 as follows:

$$P_n(x) = f(x_0) + f^1(x_0)(x - x_0) + \frac{f^2(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{f^n(x_0)}{n!}(x - x_0)^n$$

- **Theorem:** let I be an open interval and suppose that $[a, b] \subset I$. Suppose that $f \in \mathcal{C}^n(I)$ $(n \geq 0)$ and suppose that $f^{(n)}$ is differentiable on I. Then there exists $c \in (a, b)$ such that:

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

  where $P_n(x)$ denotes the Taylor polynomial of degree n at a.
- We sometimes write:

$$P_n(x) = \sum_{k\,=\,0}^{n} \frac{f^k(a)}{k!}(x - a)^k,\ and\ R_n(x) = \frac{f^{n+1}(c)}{(n+1)!}(x - a)^{n+1}$$

  Thus we can also write $f(b) = P_n(b) + R_n(b)$
- When $n = 0$ in Taylor's Theorem we get the MVT. When $n = 1$, we get the Extended Mean Value Theorem.
- Given a smooth function $f(x)$ on $a \in I \subset R$ we can write down its associated Taylor polynomials $P_n(x)$ around any point a in $\mathbb{R}$
- When we use Taylor series to approximate a function in an interval I, we must make sure that $R_n(x) \to 0$ as $n \to \infty$, for all $x \in I$
- We say that a function $f(x)$ is analytic in an (open) interval I, if for each point $a \in I$, the Taylor polynomial of the function $f(x)$ around a, converges to $f(x)$ in some (possibly smaller) interval containing a. This means that $R_n(x) \to 0$ for all x in some interval $a \in (c, d) \subset I$

Integration

- **Def:** given a closed interval $[a, b]$, a partition P of $[a, b]$ is simply a collections of points:

$$P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$$

  We can think of the points of the partition as dividing the original interval $[a, b]$ into sub-intervals $I_j = [x_{j-1}, x_j],\ 1 \geq j \geq n$

- **Def:** A partition $P' = \{a = x'_0 < x'_1 < \ldots < x'_{n-1} < x'_n = b\}$ is said to be a refinement of the partition P if for each $x_i \in P$, there exists an $x'_j \in P'$ such that $x_i = x'_j$. Intuitively, a refinement P' of a partition P will break some of the sub-intervals in P into smaller sub-intervals.

- **Def:** Given a partition $P = \{a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b\}$ and a function $f : [a,b] \to \mathbb{R}$, we define two associated quantities. First we set: $M_i = \sup_{x \in [x_{i-1}, x_i]} f(x)$ and $m_i = \inf_{x \in [x_{i-1}, x_i]} f(x),\ 1 \leq i \leq n$. We define the Lower sum as:

$$L(f,P) = \sum_{j\,=\,1}^{n} m_j (x_j - x_{j-1})$$

  Similarly, we define the Upper sum as:

$$U(f,P) = \sum_{j\,=\,1}^{n} M_j (x_j - x_{j-1})$$

- We define the lower Darboux integral of f by:

$$L(f) = \sup\{L(f,P)\ :\ P \text{ is a partition of } [a,b]\}$$

  where the supremum is taken over all partitions of $[a,b]$. Similarly, we define the upper Darboux integral of f by:

$$U(f) = \inf\{U(f,P)\ :\ P \text{ is a partition of } [a,b]\}$$

  and again the infimum is over all partitions of $[a,b]$. If $L(f) = U(f)$, then we say that f is Darboux-integrable and define:

$$\int_a^b f(t)dt = U(f) = L(f)$$

  This common value of the two integrals is called the Darboux integral.

- Properties of Darboux integral:
  - $L(f) \geq U(f)$
  - for any two partitions P_1 and P_2, we have: $L(f,P_1) \leq U(f,P_2)$
  - if P' is a refinement of P then: $L(f,P) \leq L(f,P') \leq U(f,P') \leq U(f,P)$

- Suppose that for each of the intervals I_j we are given a point $t_j \in I_j$. We will denote the collection of points t_j by t. The pair (P,t) is sometimes called a tagged partition.

- **Def:** We define the Riemann sum associated to the function f, and the tagged partition (P,t) by:

$$R(f,P,t) = \sum_{j\,=\,1}^{n} f(t_j)(x_j - x_{j-1})$$

- $$L(f,P) \leq R(f,P,t) \leq U(f,P)$$

- We define the norm of a partition P (denoted $||P||$) by $||P|| = \max_j\{|x_j - x_{j-1}|\},\ 1 \leq j \leq n$

- The Reimann integral has two definitions:
  - A function $f : [a,b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists $\delta > 0$ such that:

$$|R(f,P,t) - R| < \epsilon,$$

    whenever $||P|| < \delta$. In this case R is called the Riemann integral of the function f on the interval $[a,b]$

  - A function $f : [a, b] \to \mathbb{R}$ is said to be Riemann integrable if for some $R \in \mathbb{R}$ and every $\epsilon > 0$ there exists a $\delta > 0$ and a partition P such that for every tagged refinement (P', t') of P with $||P'|| \leq \delta$:
$$|R(f, P', t') - R| < \epsilon$$
The nice thing about the above definition is that one only has to check that $|R(f, P', t') - R|$ is small for refinements of a fixed partition, and not for all partitions

- **Theorem:** the Riemann integral (using either definition) exists if and only if the Darboux integral exists and in this case the two integrals are equal

- **Riemann integration theorem:** if $f : [a, b] \to \mathbb{R}$ be a function that is bounded, and continuous at all but finitely many points of $[a, b]$, then f is Riemann integrable on $[a, b]$. In fact, one can allow even countably many discontinuities and the theorem will remain true.

- **Theorem:** suppose f is Riemann integrable on $[a, b]$ and $c \in [a, b]$. Then:
$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt$$

- **The fundamental theorem of calculus:**
  - **Part 1:** let $f : [a, b] \to \mathbb{R}$ be a continuous function, and let:
$$F(x) = \int_a^x f(t)dt$$
for any $x \in [a, b]$. Then $F(x)$ is continuous on $[a, b]$, differentiable on (a, b) and $F'(x) = f(x)$ for all $x \in (a, b)$
  - **Part 2:** let $f : [a, b] \to \mathbb{R}$ be given and suppose there exists a continuous function $g : [a, b] \to \mathbb{R}$ which is differentiable on (a, b) and which satisfies $g'(x) = f(x)$. Then, if f is Riemann integrable on $[a, b]$, then:
$$\int_a^b f(t)dt = g(b) - g(a)$$
Note that this statement does not assume that the function $f(t)$ is continuous.

- **The mean value theorem for integration:** let $f : [a, b] \to \mathbb{R}$ be a continuous function and assume that f is differentiable in (a, b). The MVT for integration says that there exists $c \in (a, b)$ such that:
$$\int_a^b f(t)dt = f(c)(b - a)$$

Two variable functions

- The natural domain of a function is the domain on which it is defined

- The level sets of functions are the sets of the form $\{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$, where c is a constant. The level set "lives" in the xy-plane. One can also plot (in three dimensions) the surface $z = f(x, y)$. By varying the value of c in the level curves one can get a good idea of what the surface looks like. When one plots the $f(x, y) = c$ for some constant c one gets a curve. Such a curve is usually called a contour line (the contour "lives" in the $z = c$ plane).

- The graph of the function $z = x^2 + y^2$ lying above the xy-plane is a paraboloid of revolution

- The three variable definitions for limit and continuity are analogous to the two variable cases. We simply have to replace the absolute value function on $\mathbb{R}$ by the distance function on $\mathbb{R}^m$.

- **Def:** the partial derivative of $f : U \to \mathbb{R}$ with respect to x_1 at the point (a, b) is defined by:

$$\frac{\partial f}{\partial x_1}(a,b) = \lim_{x_1 \to b} \frac{f((a,x_1)) - f((a,b))}{x_1 - a}$$

  Similarly, one can define the partial derivative with respect to x_2. In this case the variable x_1 is fixed and f is regarded only as a function of x_2:

$$\frac{\partial f}{\partial x_2}(a,b) = \lim_{x_2 \to b} \frac{f((a,x_2)) - f((a,b))}{x_2 - a}$$

- Def: the partial derivatives are special cases of the directional derivative. Let $v = (v1, v2)$ be a unit vector. Then v specifies a direction in $\mathbb{R}^2$. The directional derivative of f in the direction v at a point $x = (x_1, x_2)$ is denoted by $\nabla_v f(x)$ and is defined as:

$$\lim_{t \to 0} \frac{f(x+tv) - f(x)}{t} = \lim_{t \to 0} \frac{f((x_1 + tv_1, x_2 + tv_2)) - f(x_1, x_2)}{t}$$

  $\nabla_v f(x)$ measures the rate of change of the function f at x along the path $x + tv$. If we take $v = (1, 0)$ in the above definition, we obtain $\frac{\partial f}{\partial x_1}$, while $v = (0, 1)$ yields $\frac{\partial f}{\partial x_2}$.

- All directional derivatives may exist at a point even if the function is discontinuous

- The equation of the tangent plane to $z = f(x, y)$ at the point (x_0, y_0) is:

$$f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y - y_0) = 0$$

- **Differentiability for functions of two variables:** a function $f : U \to \mathbb{R}$ is said to be differentiable at a point (x_0, y_0) if $\frac{\partial f}{\partial x}(x_0, y_0)$, and $\frac{\partial f}{\partial y}(x_0, y_0)$ exist and:

$$\lim_{(h,k) \to 0} \frac{|f(x_0 + h, y_0 + k) - f(x_0,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)h - \frac{\partial f}{\partial y}(x_0,y_0)k|}{||(h,k)||} = 0$$

  We could rewrite this as:

$$|f((x_0,y_0) + (h,k)) - f(x_0,y_0) - \frac{\partial f}{\partial x}(x_0,y_0)h - \frac{\partial f}{\partial y}(x_0,y_0)k| = \epsilon_1(h,k)||(h,k)||$$

  where $\epsilon_1(h, k)$ is a function that goes to 0 as $||(h, k)|| \to 0$. This form of differentiability now looks exactly like the one variable version.

- **Def:** we can rewrite the differentiability criterion once more as follows. We define the 1×2 matrix:

$$Df(x_0,y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0,y_0) & \frac{\partial f}{\partial y}(x_0,y_0) \end{pmatrix}$$

  The function $f(x, y)$ is said to be differentiable at a point (x_0, y_0) if there exists a matrix denoted $Df(x_0, y_0)$ with the property that:

$$f((x_0,y_0) + (h,k)) - f(x_0,y_0) - Df(x_0,y_0)\begin{pmatrix} h \\ k \end{pmatrix} = \epsilon_1(h,k)||(h,k)||$$

  for some function $\epsilon_1(h, k)$ which goes to zero as (h, k) goes to zero. Viewing the derivative as a matrix allows us to view it as a linear map from $\mathbb{R}^2 \to \mathbb{R}$. The matrix $Df(x_0, y_0)$ is called the total derivative of the function $f(x, y)$ at the point (x_0, y_0).

- When viewed as a row vector rather than as a matrix, the Derivative matrix is called the gradient and is denoted $\nabla f(x_0, y_0)$. Thus:

$$\nabla f(x_0, y_0) = (\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0))$$

  In terms of the coordinate vectors i and j the gradient can be written as:

$$\nabla f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)i + \frac{\partial f}{\partial y}(x_0, y_0)j$$

- Every differentiable function is continuous
- **Theorem:** let $f : U \to \mathbb{R}$. If the partial derivatives $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ exist and are continuous in a neighbourhood of a point (x_0, y_0) (that is in a region of the plane of the form $\{(x, y) \mid ||(x, y)-(x_0, y_0)|| < r\}$ for some $r > 0$), then f is differentiable at (x_0, y_0).
- The derivative of the composite function $z(t) = f(x(t), y(t))$ from I to $\mathbb{R}$ is given by:

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$$

n variable functions

- A continuous mapping $c : I \to \mathbb{R}^n$ of an interval I to $\mathbb{R}$ is called a curve in $\mathbb{R}^n$ $(n = 2, 3)$
- For a curve $c(t) = g(t)i + h(t)j + k(t)k$ in $\mathbb{R}^3$ its tangent or velocity vector at the point $c(t_0)$ is given by $c'(t_0) = g'(t_0)i + h'(t_0)j + k'(t_0)k$
- $$\nabla_v f = \frac{df}{dt} = \nabla f \cdot v$$
- The direction at which the function f is changing the fastest at the point (x_0, y_0, z_0):

$$v = \frac{\nabla f(x_0, y_0, z_0)}{||\nabla f(x_0, y_0, z_0)||}$$

- A general type of surface S is defined implicitly as:

$$S = \{(x, y, z) \mid f(x, y, z) = b\}$$

- If S is a surface, a tangent plane to S at a point $s \in S$ (if it exists) is a plane that contains the tangent lines at s to all curves passing through s and lying on S
- Notation f_x is for the partial derivative $\frac{\partial f}{\partial x}$
- Functions which take values in $\mathbb{R}$ are called scalar valued functions, and those functions which take values in $\mathbb{R}^n$, $n > 1$ are called vector valued functions
- Limit and continuity of n variable functions are analogous to the previous cases
- **Theorem:** let U be a subset of $\mathbb{R}^m$ $(m = 1, 2, 2, \ldots)$. The function $f : U \to \mathbb{R}^n$ is continuous if and only if each of the functions $f_i : U \to \mathbb{R},\ 1 \leq i \leq n$, is continuous.
- When m = n, vector valued functions are often called vector fields
- In physics, vector force fields that arise from scalar potential functions are called conservative fields

- **Def:** a function $f : U \to \mathbb{R}^n$, where U is a subset of $\mathbb{R}^m$ is said to be differentiable at a point x if there exists an $n \times m$ matrix $Df(x)$ such that:

$$\lim_{||h|| \to 0} \frac{||f(x+h) - f(x) - Df(x) \cdot h||}{||h||} = 0$$

  Here $x = (x_1, x_2, \ldots, x_m)$ and $h = (h_1, h_2, \ldots, h_m)$ are vectors in $\mathbb{R}^m$. The matrix $Df(x)$ is usually called the total derivative of f. It is also referred to as the Jacobian matrix.

- Properties of total derivative:

$$D(f+g)(x) = Df(x) + Dg(x)$$

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x)$$

- By the following notation:

$$\frac{\partial^n f}{\partial x_1^{n_1} \partial x_2^{n_2} \ldots \partial x_3^{n_3}}$$

  we mean to take the partial derivative of f with respect to x_k, n_k times, then take the partial derivative with respect to x_{k-1}, n_{k-1} times, and so on until you take the partial derivative with respect to x_1, n_1 times. The number n is nothing but $n_1 + n_2 + \ldots + n_k$. It is called the order of the mixed partial derivative.

- A function is said to be smooth if it belongs to $\mathcal{C}^k$ for all $k \geq 1$

- **Def:** We will say that the function $f(x, y)$ attains a local minimum at the point (x_0, y_0) (or that (x_0, y_0) is a local minimum point of f) if there is a disc:

$$D_r(x_0, y_0) = \{(x, y) \mid ||(x, y) - (x_0, y_0)|| < r\} \subseteq U$$

  of radius $r > 0$ around (x_0, y_0) such that $f(x, y) \geq f(x_0, y_0)$ for every point (x, y) in $Dr(x_0, y_0)$. Similarly, we can define a local maximum point.

- **Def:** a point (x_0, y_0) is called a critical point of $f(x, y)$ if:

$$f_x(x_0, y_0) = f_y(x_0, y_0) = 0$$

  At a critical point, the tangent plane is horizontal, that is, it is parallel to the xy-plane

- **The first derivative test:** if (x_0, y_0) is a local extremum point (that is, a minimum or a maximum point) and if $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ exist, then (x_0, y_0) is a critical point

- **Def:** the Hessian of f is defined by the matrix:

$$\begin{pmatrix} f_{xx}(x_0, y_0) & f_{xy}(x_0, y_0) \\ f_{yx}(x_0, y_0) & f_{yy}(x_0, y_0) \end{pmatrix}$$

  The determinant of the Hessian is sometimes called the discriminant and is sometimes denoted D

- Theorem: assume that (x_0, y_0) is a critical point of $f(x, y)$
  - if $D > 0$ and $f_{xx}(x_0, y_0) > 0$, then (x_0, y_0) is a local minimum of f
  - if $D > 0$ and $f_{xx}(x_0, y_0) < 0$, then (x_0, y_0) is a local maximum of f
  - if $D < 0$, then (x_0, y_0) is a saddle point of f
  - If $D = 0$, further examination of the function is necessary

- **Def:** a saddle point is a critical point which is not a local extremum (that is, a local maximum or a local minimum) of the function

- **Taylor's theorem in two varibles:** If f is a $\mathcal{C}^2$ function in a disc around (x_0, y_0), then:

$$f(x_0+h, y_0+k) = f(x_0, y_0) + f_x h + f_y k + \frac{1}{2!}[f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2] + \tilde{R}_2(h,k)$$

  where $\tilde{R}_2(h,k)/||(h,k)||^2 \to 0$ as $||(h,k)|| \to 0$

- Closed bounded intervals are called compact sets

- **Theorem:** a continuous function on a compact set in $\mathbb{R}^2$ will attain its extreme values

- **Def:** a point (x_0, y_0) such that $f(x,y) \leq f(x_0,y_0)$ or $f(x,y) \geq f(x_0,y_0)$ for all (x,y) in the domain being considered is called a global maximum or minimum point respectively

- Suppose we are given a function $f(x,y)$ in two variables. We would like maximize or minimize it subject to the constraint that $g(x,y) = c$. In geometric terms, we want to find the maximum or minimum values of f while staying on the curve $g(x,y) = c$. Then we are looking for points (x_0, y_0) such that:

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0)$$

  subject to the constraint condition, $g(x_0, y_0) = c$. The λ above is called the Lagrange multiplier.

- **The four squares theorem:** every positive integer can be written as a sum of four squares

- **Theorem:** the function $x_1^2 + x_2^2 + x_3^2 + x_4^2$ represents every natural number

- An n-ary quadratic form over the real numbers is a function from $\mathbb{R}^n$ or $\mathbb{Z}^n$ to $\mathbb{R}$ of the form:

$$q(x_1, x_2, \ldots, x_n) = \sum_{1 \,\leq\, i,\, j \,\leq\, n} q_{ij}\, x_i\, x_j, \;\; a_{ij} \in \mathbb{R}$$

  The example $x_1^2 + x_2^2 + x_3^2 + x_4^2$ is an example of a quartenary quadratic form. It is a diagonal form, that is, only square terms appear.

- A quadratic form is called positive definite if $q(x_1, \ldots, x_n) > 0$ for all $(x_1, \ldots, x_n)$ in $\mathbb{R}^n \backslash \{(0, 0, \ldots, 0)\}$

- **The Bhargava-Hanke theorem:** if a positive definite (integral) quadratic form represents every number $n \leq 290$, then it represents all natural numbers

- Any rectangle R in the plane can be described as the set of points in the cartesian product $[a,b] \times [c,d]$ of two closed intervals

- For taking a partition of the above rectangle we take a partition P_1 of $[a,b]$ and a partition P_2 of $[c,d]$ and take the product of the two partitions. Thus if $P_1 = \{a = x_0, x_1, \ldots, x_m = b\}$ and $P_2 = \{c = y_0, y_1, \ldots, y_n = d\}$, we take the collection of points $P = \{(x_i, y_j) \mid 1 \leq i \leq m,\ 1 \leq j \leq n$. The point (x_i, y_j) is the left bottom corner of the rectangle $R_{ij} = (x_i, x_{i+1}) \times (y_i, y_{j+1})$. As i and j vary, we get a family of rectangles $R_{ij},\ 0 \leq i \leq m-1,\ 0 \leq j \leq n-1$. By identifying each rectangle with its left bottom corner we can think of P as the collection of these rectangles R_{ij}. Clearly, $R = \cup_{i,j} R_{ij}$, and the collection of rectangles P is called a partition of R.