Part I

ME010202 Advanced Topology

Chapter 7

Separation Axioms

7.1 Compactness and Separation Axioms

Proposition 7.1. *Let X be a T_2 space, $x \in X$ and F is a compact subset of X not containing x. Then there exist opensets U, V such that $x \in U$, $F \subset V$ & $U \cap V = \phi$.*

Corollary 7.1.1. *A compact subset in a T_2 space is closed.*

Corollary 7.1.2. *Every map from a compact space into a T_2 space is closed. And its range is a quotient space of the domain.*

Corollary 7.1.3. *A continuous bijection from a compact space onto a T_2 sspace is a homeomorphism.*

Corollary 7.1.4. *Every continuous, one-to-one function from a compact space into a T_2 space is an embedding.*

Theorem 7.2. *Every compact T_2 space is a T_3 space.*

Proposition 7.3. *Let X be a regular space, C a closed subset of X and F a compact subset of X, such that $C \cap F = \phi$. Then there exist open sets U, V such that $C \subset U$, $F \subset V$ and $U \cap V = \phi$.*

Theorem 7.4. *Every regular, Lindeloff space is normal.*

Corollary 7.4.1. *Every regular, second countable space is normal.*

Corollary 7.4.2. *Every compact T_2 space is T_4.*

7.2 The Urysohn Characterisation of Normality

Proposition 7.5. *Let A B be subsets of a space X and suppose there exists a continuous function $f : X \to [0,1]$, such that $f(x) = 0$, $\forall x \in A$ and $f(x) = 1$, $\forall x \in B$. Then there exists disjoint open sets U, V such that $A \subset U$ and $B \subset V$.*

Corollary 7.5.1. *If X has the property that for any disjoint closed subsets A, B of X, there exists a continuous function $f : X \to [0,1]$ such that $f(x) = 0$, $\forall x \in A$ and $f(x) = 1$, $\forall x \in B$, then X is normal.*

Theorem 7.6. *A topological space X is normal iff it has the property that for every mutually disjoint, closed subsets A, B of X, there exists a continuous function $f : X \to [0,1]$ such that $f(x) = 0$ for all $x \in A$ and $f(x) = 1$ for all $x \in B$*

Lemma 7.7. *Let $f : X \to [0,1]$ be continuous. For each $t \in \mathbb{R}$ let $F_t\{x \in X : f(x) < t\}$. Then the indexed family $\{F_t : t \in \mathbb{R}\}$ has the following properties*

1. *F_t is an open subset of X for each $t \in \mathbb{R}$*
2. *$F_t = \phi$ for $t < 0$*
3. *$F_t = X$ for $t > 1$*
4. *For any $s,\ t \in \mathbb{R},\ s < t \implies \overline{F_s} \subset F_t$.*

Moreover, for each $x \in X,\ f(x) = \inf\{t \in \mathbb{Q} : x \in F_t\}$.

Lemma 7.8. *Let X be a topological space and suppose $\{F_t : t \in \mathbb{Q}\}$ is a family of sets in X such that*

1. *F_t is open in X for each $t \in \mathbb{Q}$*
2. *$F_t = \phi$ for $t \in \mathbb{Q},\ t < 0$*
3. *$F_t = X$ for $t \in \mathbb{Q},\ t > 1$*
4. *$\overline{F_s} \subset F_t$ for $s,\ t \in \mathbb{Q},\ s < t$*

For $x \in X$, let $f(x) = \inf\{t \in \mathbb{Q} : x \in F_t\}$. Then f is a continuous real-valued function on X and it takes values in the unit interval $[0,1]$.

Corollary 7.8.1. *All T_4 spaces are completely regular and hence Tychonoff.*

Proof. Let $x \in X$ and D be closed subset not containing x. We have X is a T_4 space. Therefore X is T_1 as well as normal. Now the singleton set, $\{x\}$ is closed, since X is a T_1 space. And by Urysohn's lemma for disjoint, closed subsets $\{x\},\ D$ there exists a continuous, real-valued function $f : X \to [0,1]$ such that $f(x) = 0$ and $f(y) = 1$ for all $y \in D$. Therefore the space X is completely regular and hence Tychonoff. □

Remark (Urysohn function)**.** *The function whose existence is asserted by Urysohn's lemma is called a Urysohn function*

7.3 Tietze Characterisation of Normality

Proposition 7.9. *Let A be a subset of a space X and let $f : A \to \mathbb{R}$ be continuous. Then any two extensions of f to X agree on $\overline{A}$. In other words, if at all an extension of f exists its values on $\overline{A}$ are uniquely determined by values of f on A.*

Proposition 7.10. *Suppose a topological space X has the property that for every closed subset A of X, every continuous real valued function on A has a continuous extension to X. Then X is normal.*

Definitions 7.11 (Pointwise Convergence)**.** *Let X be a topological space and (Y, d) a metric space. Then a sequence of functions $\{f_n\}$ from X to Y converges pointwise to f if for every $x \in X$ the sequence $\{f_n(x)\}$ converges to $f(x)$ in Y.*

In other words, given a very small value, $\epsilon > 0$, there exists some $\delta > 0$ such that for every $x \in X$ there exists $N_x \in \mathbb{N}$. This N_x may be different for different values of x and for every $n > N_x$, $d(f(x), f_n(x)) < \delta$.

Definitions 7.12 (Uniform Convergence)**.** *Let X be a topological space and (Y, d) a metric space. Then a sequence of functions $\{f_n\}$ from X to Y converges uniformly to f if given a small $\epsilon > 0$, there exists $\delta > 0$ such that there exists $N \in \mathbb{N}$. This N is independent of the value of x and for every $n > N$, $d(f(x), f_n(x)) < \delta$.*

Proposition 7.13. *Let X, (Y, d), $\{f_n\}$ and f be as above and suppose $\{f_n\}$ converges to f uniformly. If each f_n is continuous, then f is continuous.*

Definitions 7.14 (Uniform Convergence of Series)**.** *Let X be a topological space and (Y, d) be a metric space. Then a series of function $\sum_{n=1}^{\infty} f_n$ converges uniformly to f if the sequence of partial sums converges uniformly to f.*

In other words, let $g_m = \sum_{n=1}^{m} f_n$. Then $\sum_{n=1}^{\infty} f_n$ converges to f uniformly if the sequence of partial sums $\{g_m\}$ converges to f uniformly.

Proposition 7.15. *Let $\sum_{n=1}^{\infty} M_n$ be a convergent series of non-negative real numbers. Suppose $\{f_n\}$ is a sequence of real valued functions on a space X such that for each $x \in X$ and $n \in \mathbb{N}$, $|f_n(x)| \leq M_n$. Then the series $\sum_{n=1}^{\infty} f_n$ converges uniformly to a real valued function on X.*

—continue page 185—

Chapter 8

Products and Coproducts

8.1 Cartesian Products of Families of Sets

8.2 The Product Topology

8.3 Productive Properties

Chapter 9

Embedding and Metrisation

Chapter 10

Nets and Filters

10.1 Definition and Convergence of Nets

Definitions 10.1 (Directed Set)**.** *[Joshi, 1983, 10.1.1]*
A directed set D is a pair $(D, \geq)$ where D is a nonempty set and ge is a binary relation on D such that

1. *The relation 'follows'($\geq$) is transitive. ie, $m \geq n,\ n \geq p \implies m \geq p$*
2. *The relation 'follows'($\geq$) is reflexive. ie, For every $m \in D,\ m \geq m$*
3. *For any $m, n \in D$, there exists $p \in D$ such that $p \geq m$ and $p \geq n$.*

sequence in a set X is a function f from the set of all integers into X.

Definitions 10.2 (Net)**.** *[Joshi, 1983, 10.1.2]*
A net in a set X is a function S from a directed set D into the set X.

Remark. *The set $\mathbb{N}$ together with the relation 'less than or equal to'($\leq$) is a directed set. Clearly, the relation 'less than or equal to' is reflexive and trasitive. And the third condition is true iff every finite subset E of D has an element $p \in E$ such that p follows each element of E. This is a weaker notion compared to the well ordering principle[1] of the set of all integers. Thus $\mathbb{N}$ is a directed set and every sequence in X is also a net in X.*

Remark (Significance of Net)**.** *A net on a set is a generalisation of 'a sequence on a set' obtained by simplifying the domain of the sequence into a directed set. The notion directed set is derived by assuming a few properties of $\mathbb{N}$.*

The convergence of sequence is not strong enough to characterise topologies as the limit of convergent sequences are unique for both Hausdorff and Co-countable spaces. The notion of Net allows us to differentiate between Hausdorff spaces from Co-countable spaces in terms of convergence of nets. The limit of a convergent net on a topological space is unique iff it is a Hausdorff space. ie, We have removed a few restrictions, so that we will have some convergent nets (which are obviously not sequences) with multiple limit points for Co-countable spaces.

[1] Well-ordering principle : Every subset of $\mathbb{N}$ has a least element in it.

Remark. *Examples of Directed Sets*

1. *Let X be a topological space and $x \in X$. Then the neighbourhood system $\mathcal{N}_x$ is a directed set with the binary relation $\subset$ (subset/inclusion).*

   (a) *Let $U,\ V,\ W$ be any three neighbourhoods of $x \in X$ such that $U \subset V$ and $V \subset W$. Then, clearly $U \subset W$.*
   Therefore, $U \geq V,\ V \geq W \implies U \geq W$.

   (b) *Let U be any neighbourhood of $x \in X$, then $U \subset U$.*
   Therefore, $U \geq U$.

   (c) *Let U, V be any two neighbourhoods of $x \in X$, then there exists their intersection $W = U \cap V$, which is a neighbourhood of x. Clearly $W \subset U$ and $W \subset V$.*
   Therefore $\forall U, V \in \mathcal{N}_x, \exists W \in \mathcal{N}_x$ such that $W \geq U$ and $W \geq V$.

2. *Let $\mathcal{P}$ be the set of all partitions on closed unit interval $[0, 1]$. A partition $P \in \mathcal{P}$ is a refinement of $Q \in \mathcal{P}$ if every subinterval in P is contained in some subinterval of Q. Then $\mathcal{P}$ with the binary relation refinement is a directed set.*

   (a) *Suppose $P,\ Q,\ R$ are three partitions of $[0, 1]$ such that P is a refinement of Q and Q is a refinement of R, then clearly P is a refinement of R since each subinterval of P is contained some subinterval of Q, which is contained in some subinterval of R.*
   Therefore, $P \geq Q,\ Q \geq R \implies P \geq R$

   (b) *Suppose P is a partition of $[0, 1]$. Then trivialy, P is a refinement of itself since every subinterval of P is contained in the same subinterval of P.*
   Therefore, $\forall P \in \mathcal{P},\ P \geq P$

   (c) *Suppose $P,\ Q$ be any two partition of $[0, 1]$. Then $R = P \cup Q$ is a refinement of both the partitions.*
   Therefore $\forall P, Q \in \mathcal{P},\ \exists R \in \mathcal{P}$ such that $R \geq P$ and $R \geq Q$

For example, let $P = \{0,\ 0.3,\ 0.7,\ 1\}$. Then the subintervals in P are $[0, 0.3]$, $[0.3, 0.7]$ and $[0.7, 1]$. Let $Q = \{0,\ 0.3,\ 0.5,\ 1\}$ and $R = \{0,\ 0.3,\ 0.5,\ 0.7,\ 1\}$. Then R is a refinement of P, but Q is not a refinement of P since there is a subinterval $[0.5, 1]$ in Q which is not properly contained in any subinterval of P. However, R is a refinement of Q as well.

Remark. *Examples of Nets*

1. *Let X be a topological space and $x \in X$. Let $\mathcal{N}_x$ be the set of all neighbourhoods of x. Let $D = (\mathcal{N}_x, X)$ be the directed set given by $(N, y) \in (\mathcal{N}_x, X)$ if $N \in \mathcal{N}_x$ and $y \in N$ and $(N, y) \geq (M, z)$ if $N \subset M$. Then the function $S : (\mathcal{N}_x, X) \to X$ given by $S(N, y) = y$ is a net on X.*

   For example, let $X = \{a, b, c, d\}$ and $\mathcal{T} = \{\{a\},\ \{a, b\},\ \{a, b, c\},\ \{a, b, c, d\}\}$. Also let $S : (\mathcal{N}_b, X) \to X$ defined by $S(N, y) = y$. Suppose $C = \{a, b, c\}$. Then $C \in \mathcal{N}_b$. ie, C is a neighbourhood of b. Then $S(C, c) = c$.

2. *Riemann Net - Let $D = (\mathcal{P}, \xi)$ where $\mathcal{P}$ is the set of all partitions on $[0,1]$ and ξ is a finite sequence in $[0,1]$ such that consecutive terms belongs to consecutive subintervals of the partition. The set $(\mathcal{P}, \xi)$ is directed set with $\geq$ given by $(P, \eta) \geq (Q, \psi)$ iff P is a refinement of Q.*

   For example, let $P \in \mathcal{P}$ is given by $P = \{\ 0,\ 0.3,\ 0.7,\ 1\ \}$ and $\eta = \{\ 0.2,\ 0.6,\ 0.9\ \}$. Then $(P, \eta) \in (\mathcal{P}, \xi)$.

   Let $f : \mathbb{R} \to \mathbb{R}$ be any function, then the function,

$$S : (\mathcal{P}, \xi) \to \mathbb{R} \text{ defined by } S(P, \eta) = \sum_{j=1}^{k} f(\eta_k)(a_k - a_{k-1})$$

   where $P = \{a_0,\ a_1, \cdots,\ a_k\ \}$ is the Riemann Net with respect to the real function f.

   For example, let $f(x) = x^2$ and P, η are same as above example, then $S(P, \eta) = 0.2^2(0.3 - 0) + 0.6^2(0.7 - 0.3) + 0.9(1 - 0.7) = 3.99$

Definitions 10.3 (Convergence of a Net)**.** *[Joshi, 1983, 10.1.3]*
A net $S : D \to X$ converges to a point $x \in X$ if for any nbd U of x, there exists $m \in D$ such that $n \in D,\ n \geq m \implies S(n) \in U$. And x is a limit of the net S.

Remark. *The choice of m depends on the choice of neighbourhood U.*

$$S : D \to X,\ S \to x \iff (\forall U \in \mathcal{N}_x,\ \exists m_U \in D,\ \text{such that } n \geq m_U \implies S(n) \in U)$$

Theorem 10.4 (Net characterisation of Hausdorff space)**.** *[Joshi, 1983, 10.1.4]*
A topological space is Hausdorff iff limits of all nets in it are unique.

Proof. Let X be a Hausdorff space. Suppose $S : D \to X$ is net on X such that S converges to two distinct points $x, y \in X$. Since X is a Hausdorff space and $x \neq y$, there exists open sets U, V such that $x \in U,\ y \in V, U \cap V = \phi$.

The net S converges to $x \in X$, therefore $\exists m_x \in D$ such that $n \geq m_x \implies S(n) \in U$ And, the net S converges to $y \in X$, therefore $\exists m_y \in D$ such that $n \geq m_y \implies S(n) \in V$.

Since D is a directed set and $m_x, m_y \in D$, there exists $p \in D$ such that $p \geq m_x$ and $p \geq m_y$. Now, $n \geq p \implies n \geq m_x,\ n \geq m_y$, since $\geq$ is transitive. (ie, $n \geq p,\ p \geq m_x \implies n \geq m_x$, and $n \geq p,\ p \geq m_y \implies n \geq m_y$).

We have $n \geq p \implies n \geq m_x$ and $n \geq m_x \implies S(n) \in U$. Therefore, $n \geq p \implies S(n) \in U$. Similarly, $n \geq p \implies n \geq m_y \implies S(n) \in V$. Therefore $S(n) \in U \cap V$ which is a contradiction, since $U \cap V = \phi$. Therefore, if a net S converges to two points x, y, then $x = y$. That is, if a net S in a Hausdorff space X is convergent then its limit is unique.

Conversely, suppose that X is a topological space and every convergent net in X has a unique limit. Suppose X is not a Haudorff space. Then there exists at least two distinct points $x, y \in X$ such that every neighbourhood of x intersects with every neighbourhood of y. Now consider the set $D = \mathcal{N}_x \times \mathcal{N}_y$ and relation $\geq$ on D such that $(U_1, V_1) \geq (U_2, V_2)$ if $U_1 \subset U_2$ and $V_1 \subset V_2$.

By the axiom of choice, a function $S : D \to X$ such that $S(U, V) \in U \cap V$ is well defined, since every nbd of x intersects every nbd of y. Thus, S is a net in X. We claim that S converges to both x and y.

Let U be a nbd of x. Then $S(U', V') \in U' \cap V'$. We have $(U, X) \in D$ such that $(U', V') \geq (U, X) \implies U' \subset U$. Then, $S(U', V') \in U' \cap V' \subset U \cap X = U$. Thus, for any nbd U containing x, we have $(U, X) \in D$ such that $(U', V') \geq (U, X) \implies S(U', V') \in U$. Therefore, S converges to $x \in X$.

Similarly, Let V be a nbd of y. Then for any nbd V containing y, we have $(X, V) \in D$ such that $(U', V') \geq (X, V) \implies S(U', V') \in V$, since $S(U', V') \in U' \cap V' \subset X \cap V = V$. Therefore, S converges to $y \in X$ as well, where $x \neq y$. This is a contradition to the assumption that every convergent net in X has a unique limit. Therefore, for any two points $x, y \in X$, there should be some nbd of x that doesn't intersect some nbd of y. Therefore, X is a Hausdorff space. □

Definitions 10.5 (Eventual Subset)**.** *[Joshi, 1983, 10.1.5]*
A subset E of a directed set D is an eventual subset of D if there exists $m \in D$ such that $n \geq m \implies n \in E$.

Remark. *Let E be an eventual subset of D such that $n \geq m \implies n \in E$. Then $p \in E \not\implies p \geq m$. ie, Subset E may contain elements that doesn't follow the above m.*

Remark. *[Joshi, 1983, 10.1.6]*
Let E be an eventual subset of D, then E is a directed set.

1. $m, n, p \in E,\ m \geq n,\ n \geq p \implies m, n, p \in D,\ m \geq n,\ n \geq p \implies m \geq p$
2. $m \in E \implies m \in D \implies m \geq m$
3. $m, n \in E \implies m, n \in D \implies \exists p \in D$ *such that* $p \geq m$ *and* $\geq n$.

   $\exists m' \in D$ *such that* $n' \geq m' \implies n' \in E$. *(E is an eventual subset of D)*

   $p, m' \in D \implies \exists p' \in D$ *such that* $p' \geq p$ *and* $p' \geq m'$*.(D is a directed Set)*

   $p' \geq m' \implies p' \in E$. *(E is eventual subset of D with respect to m')*

   $p' \geq p,\ p \geq m \implies p' \geq m$ *and* $p' \geq p,\ p \geq n \implies p' \geq n$.

   Therefore $\forall m, n \in E,\ \exists p' \in E$ *such that* $p' \geq m$ *and* $p' \geq n$.

Definitions 10.6 (Net eventually in A)**.** *[Joshi, 1983, 10.1.5]*
Let $S : D \to X$ be a net in a topological space X. Then S is eventually in subset A of X if $S^{-1}(A)$ is an eventual subset of D.

Remark. *Let $S : D \to X$ be a net in X. Then S converges to $x \in X$ if S is eventually in each nbd U of x.*

Definitions 10.7 (Cofinal subset)**.** *[Joshi, 1983, 10.1.7]*
A subset F of a directed D is a cofinal subset of D if for any $m \in D$, there exists $n \in F$ such that $n \geq m$.

Remark. *Let X be a topological space and $x \in X$. Let $\mathcal{N}_x$ be the set of all neighbourhood of x and $\mathcal{L}$ be a local base of X at x. We have, $(\mathcal{N}_x, \geq)$ is a directed set where $\forall U, V \in \mathcal{N}_x,\ U \geq V \iff U \subset V$, then $\mathcal{L}$ is cofinal in $\mathcal{N}_x$.*

Remark. *[Joshi, 1983, 10.1.8]*
Let F be a cofinal subset of D, then F is a directed set.

1. $m, n, p \in F,\ m \geq n,\ n \geq p \implies m, n, p \in D,\ m \geq n,\ n \geq p \implies m \geq p$
2. $m \in F \implies m \in D \implies m \geq m$
3. $m, n \in F \implies m, n \in D \implies \exists p \in D$ *such that* $p \geq m$ *and* $\geq n$.

   E *is cofinal,* $p \in D \implies \exists p' \in F$ *such that* $p' \geq p$.

   $p' \geq p,\ p \geq m \implies p' \geq m$ *and* $p' \geq p,\ p \geq n \implies p' \geq n$.

   Therefore $\forall m, n \in F,\ \exists p' \in F$ *such that* $p' \geq m$ *and* $p' \geq n$.

Definitions 10.8 (Net frequently in A)**.** *[Joshi, 1983, 10.1.7]*
Let $S : D \to X$ be a net in a topological space X. Then S is frequently in subset B of X if $S^{-1}(B)$ is a cofinal subset of D.

Proposition 10.9. *[Joshi, 1983, 10.1.6]*
Let $S : D \to X$ be a net in a topological space X. Let E be an eventual subset of D. Then, S converges to x iff $S_{/_E}$ converges to x.[Joshi, 1983, 10.1.6]

Proof. Let $S : D \to X$ be a net in X, E be an eventual subset of D, and $x \in X$. Then, $S_{/_E} : E \to X$ is defined by $n \in E \implies S_{/_E}(n) = S(n)$

Suppose S converges to x. Let U be a nbd of x, then S is eventually in U. ie, $S^{-1}(U)$ is an eventual subset of D. Then $\exists m \in D$ such that $n \geq m \implies n \in S^{-1}(U) \implies S(n) \in U$. Since set E is eventual subset of D, $\exists m' \in D$ such that $n \geq m' \implies n \in E$.

Since E is a directed set, $S_{/_E} : E \to X$ is a net in X. And $m, m' \in D \implies \exists p \in D$ such that $p \geq m$ and $p \geq m'$. We have, $p \geq m' \implies p \in E$. And $n \geq' p \implies n \geq p,\ p \geq m \implies n \geq m \implies S(n) \in U \implies S_{/_E}(n) \in U$. Therefore, $n \geq' p \implies S_{/_E}(n) \in U$. Since U is arbitrary, $S_{/_E}$ converges to x.

Conversely, suppose that $S_{/_E}$ converges to x. Let U be a nbd of x, then $S_{/_E}$ is eventually in U. ie, $S^{-1}_{/_E}(U)$ is an eventual subset of D. ie, $\exists m \in D$ such

that $n \geq m \implies n \in S^{-1}_{/_E}(U) \implies S_{/_E}(n) \in U \implies S(n) \in U$. Therefore, $n \geq m \implies S(n) \in U$. Since, U is arbitrary, S converges to every nbd of x. ie, S converges to x. □

Proposition 10.10. *[Joshi, 1983, 10.1.8]*
Let $S : D \to X$ be a net in a topological space X. Let F be a cofinal subset of D. If S converges to x, then $S_{/_F}$ converges to x.

Proof. Let $S : D \to X$ be a net in X and S converges to $x \in X$. Also let F be a cofinal subset of D. Then $S_{/_F}$ is also a net in X, since $(F, \geq')$ is a directed set where $\forall m, n \in F,\ m \geq n \implies m \geq' n$.

Since S converges to x, for any nbd U of x, $\exists m \in D$, such that $n \geq m \implies S(n) \in U$. Since F is cofinal, $\exists p \in F$ such that $p \geq m$. Thus $n \geq' p \implies n \geq p,\ p \geq m \implies n \geq m \implies S(n) \in U \implies S_{/_F}(n) \in U$. Therefore, $\exists p \in F$ such that $n \geq' p \implies S_{/_F}(n) \in U$. Since U is arbitrary, $S_{/_F}$ is eventually in every nbd of x. ie, $S_{/_F}$ converges to x. □

Remark. *But converse of the above is not true. $S_{/_F}$ converges to x does not imply that S converges to x, since cofinal subset F not necessarily contain every element following a particular m.*

Definitions 10.11 (Cluster point)**.** *[Joshi, 1983, 10.1.9]*
Let $S : D \to X$ be a net in a topological space X. Then $x \in X$ is a cluster point of S, if S is frequently in each nbd U of x in X.

Proposition 10.12. *[Joshi, 1983, 10.1.10]*
Let $S : D \to X$ be a net in a topological space X. Then $x \in X$ is a cluster points of X, if $S_{/_F}$ converges to x for some cofinal subset F of D.

Proof. Let $S : D \to X$ be a net in X and $(F, \geq')$ be a cofinal subset of $(D, \geq)$. Then $S_{/_F}$ is also a net in X. Suppose $S_{/_F}$ converges to $x \in X$. Let U be a nbd of x, then $\exists m \in F$ such that $n \geq' m \implies S_{/_F}(n) \in U$.

Let $m' \in D$. Then $\exists p' \in F$ such that $p' \geq m'$, since F is a cofinal subset of D. We have, $m, p' \in F$, then $\exists p \in F$ such that $p \geq' m$ and $p \geq' p'$. Since $F \subset D$, we have $p, m \in F \implies p, m \in D$ and $p \geq' m \implies p \geq m$.

Also $p \geq' m \implies S_{/_F}(p) \in U \implies S(p) \in U$. Therefore, $\forall m' \in D,\ \exists p \in D$ such that $p \geq m'$ and $S(p) \in U$. Since U, m' are arbitrary, S is frequently in every nbd of x. ie, x is a cluster point of S. □

Definitions 10.13 (Subnet)**.** *[Joshi, 1983, 10.1.11]*
Let $S : D \to X$ be a net in a topological space X. Then a net $T : E \to X$ in X, is a subnet of S if there exists a function $N : E \to D$ such that $S \circ N = T$ and $\forall m \in D,\ \exists p \in E$ such that $n \geq' p \implies N(n) \geq m$.

Remark. *A net $T : E \to X$ is a subnet of $S : D \to X$ if $\exists N : E \to D$ such that $S \circ N = T$ and S is frequently in $T(E)$.*
Let $T : E \to X$ be a subnet of $S : D \to X$ and A be a subset of X. If T eventually in A, then S is frequently in A.

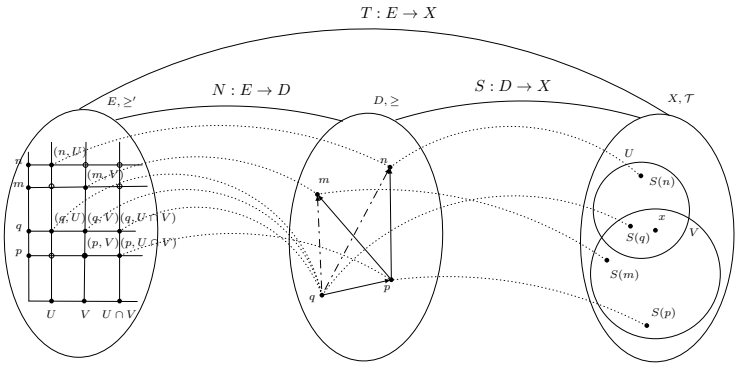

Figure 10.1: $\forall (n, U), (m, V) \in E,\ \exists (q, W) \in E$ such that $(q, W) \geq' (n, U),\ (q, W) \geq' (m, V)$

Proposition 10.14. *[Joshi, 1983, 10.1.12]*
Let $S : D \to X$ be a net in a topological space X. Then $x \in X$ is a cluster point of S iff there exists a subnet of S which converges to x.

Synopsis. *Let $(D, \geq)$, $(E, \geq')$ be two directed sets. And $T : E \to X$ be a subnet of $S : D \to X$.*

If T converges to x, then T is eventually in each nbd U of x. And since T is a subnet of S, there exists $N : E \to D$ such that $N(E)$ is a cofinal subset of D. Therefore, S is frequently in each nbd U of x. Thus, x is a cluster point of S.

If x is a cluster point of X, then S is frequently in every nbd of x. Let $N : E \to D$ be $N(n, U) = n$. Construct a directed subset E of $D \times \mathcal{N}_x$ such that $(n, U) \in E \iff S(n) \in U$. Now T is eventually in every nbd U of x, as those points with images outside U are removed by construction. Therefore, it is sufficient to show that $T : E \to X$ is a subnet of the net $S : D \to X$. Clearly, E is a directed set and $N : E \to D$ defined by $N(n, U) = n$ satisfies both $S \circ N = T$ and $\forall n \in D$, $\exists p \in E$ such that $m \geq p \implies N(m) \geq n$.

Proof. Let $S : D \to X$ be a net in X. Suppose there exists a subnet $T : E \to X$ that converges to $x \in X$. By the definition of subnet, we have $\exists N : E \to D$ such that $S \circ N = T$ and S is frequently in $T(E)$.

We have, T convergers to x, thus for any neighbourhood U of x, there exists $m' \in E$ such that $n' \geq' m' \implies T(n') \in U$.

Also we have, T is a subnet of S. Then $\exists N : E \to D$ such that $\forall m \in D$, $\exists p' \in E$ such that $n' \geq' p' \implies N(n') \geq m$.

Now, for any $m \in D$, we have $m', p' \in E$. Since E is a directed set, there exists $n' \in E$ such that $n' \geq' m'$ and $n' \geq' p'$.

Then, $n' \geq m' \implies T(n') \in U$ and $n' \geq' p' \implies N(n') \geq m$.

Thus for any $m \in D$, there exists $N(n') = n \in D$ such that $S(n) = S(N(n')) = T(n') \in U$.

Thus S is frequently in any neighbourhood U of x. Therefore, x is a cluster point of S.

Conversely, suppose that x is a cluster point of S. We have to construct a directed set $(E, \geq')$ and a function $N : E \to D$ such that T is a subnet of S and T converges to x. Let $\mathcal{N}_x$ be the family of all neighbourhood of x in X.

Consider $E = \{(n, U) \in D \times \mathcal{N}(x) : S(n) \in U\}$ and define $\geq'$ by $(n, U) \geq' (m, V)$ if $n \geq m$ and $U \subset V$. Trivially, $(n, U) \geq' (m, V) \geq' (p, W) \implies (n, U) \geq' (p, W)$ and $(n, U) \geq' (n, U)$. Also, for any $(n, U), (m, V) \in E$, we have $n, m \in D$ and $U, V \in \mathcal{N}(x)$. Since D is a directed set, $\exists p \in D$ such that $p \geq n$ and $p \geq m$. Also, $U \cap V \in \mathcal{N}_x$ and there exists $q \in D$ such that $S(q) \in U \cap V$ and $q \geq' p$, since S is frequently in every nbd of x. And $U \cap V \in \mathcal{N}(x)$ such that $U \cap V \subset U$ and $U \cap V \subset V$. Thus $\exists (q, U \cap V) \in E$ such that $(q, U \cap V) \geq' (n, U)$ and $(q, U \cap V) \geq' (m, V)$. Therefore, $(E, \geq')$ is a directed set.

Define $N : E \to D$ by $N(n, U) = n$. Again for any $(m, V) \in E$, there exists $m \in D$ such that $(n, U) \geq' (m, U)$ implies there exists $n \in D$ such that $N(n, U) = n$ and $n \geq m$.

It remains to prove that, T converges to x. Let $U \in \mathcal{N}_x$ be a nbd of x in X. We have, x is a cluster points of S. Therefore, $\forall m \in D$, $\exists p \in D$ such that $S(p) \in U$. By the constructon of E, we have $(p, U) \in E$. Suppose $(n, V) \geq' (p, U)$, then $n \geq p$, $V \subset U$, and $S(n) \in V$. Clearly $S(n) \in U$. Therefore, $\forall (n, V) \geq' (p, U)$, $T(n, V) \in U$. That is, T is eventually in every nbd of x. ie, subnet T is convergent to x. Therefore, for each cluster point of the net S, there exists some subset converging to it. □

Remark.

- *Importance of Construction of E*
  If x is a cluster point of a net S in X, then S is frequently in some cofinal subset of D. Thus, if we consider any cofinal subset D' of D which is a direct set with $\geq$ restricted to D'. Then $N : D' \to D$ defined by $N(n) = n$ gives a subnet $T : D' \to X$ of the net S. However, this subnet need not converge to x. The strongest statement, we can make on T is that 'x is a cluster point of T', since $N : D \times \mathcal{N}(x) \to D$, $N(n) = n$ is completely independent of U. This problem is overcame by constructing E which is dependent on each nbd U of x.

- *Existence of $q \in D$ such that $q \geq' p$ and $S(q) \in U$. We have, p follows both n&m and $U \cap V$ is a subset of both U&V. However, since S is only frequently in U, p not necessarily be in U. But there is always someone following p which has its image in U. This q follows both n&m, since $\geq'$ is transitive.*

Chapter 11

Compactness

11.1 Variations of Compactness

In this chapter, we have two other notions of compactness - countable compactness and sequential compactness.[1]

Compact A topological space is compact iff every open cover of it has a finite subcover. ([Joshi, 1983, 6.1.1]) [Heine-Borel]

Countably Compact A topological space is countably compact iff every countable, open cover of it has a finite subcover. [Joshi, 1983, 11.1.1]

Sequentially Compact A topological space is sequentially compact iff every sequence in it has a convergent subsequence. [Joshi, 1983, 11.1.8] [Bolzano-Weierstrass]

Countable compactness is a weaker notion compared to compactness.[2] However, sequentially compact and compact are not necessarily comparable.[3].

We have seen earlier that compactness has the following properties 1. compactness is weakly hereditary.[Joshi, 1983, 6.1.10] 2. compactness is preserved under continuous functions.[Joshi, 1983, 6.1.8] 3. every continuous real functions on compact space is bounded and attains its extrema.[Joshi, 1983, 6.1.6] 4. every continuous real function on a compact, metric space is uniformly continuous by Lebesgue covering lemma.[Joshi, 1983, 6.1.7]

Countably compact spaces, Sequentially compact spaces have all the four properites listed above.

11.1.1 Countable compactness

Weakly hereditary property

A subspace $(A, \mathcal{T}_{/A})$ being countably compact doesn't imply that $(X, \mathcal{T})$ is countably compact. However, if $(X, \mathcal{T})$ is a countably compact space and A

[1] For $\mathbb{R}$, Compactness & Sequentially compactness are equivalent to the completeness axiom.
[2] Every compact space is countably compact.
[3] $\mathcal{T}_1, \mathcal{T}_2$ are non-comparable, if $\mathcal{T}_1 \not\subset \mathcal{T}_2$ and $\mathcal{T}_2 \not\subset \mathcal{T}_1$.[Joshi, 1983, 4.2.1]

is a closed subset of X, then $(A, \mathcal{T}_{/A})$ is also a countably compact space. In other words, countably compactness is weakly hereditary.

Theorem 11.1. *Countable compactness is weakly hereditary.[Joshi, 1983, 11.1.3]*

Synopsis. *Let A be a closed subset of countably compact space, X. If A has a countable open cover $\mathcal{U}$, then we can obtain a respective countable, open cover for X by attaching $X - A$ to the extensions of members of $\mathcal{U}$ to X. This cover has a finite subcover. Then restricting them to A, we get a finite subcover of $\mathcal{U}$.*

Proof. Suppose X is a countably compact space. And A is a closed subset of X. We need to show that A is countably compact. Without loss of generality,[4] assume that A is a proper subset of X. Then $X - A$ is a non-empty, open subset of X.

Let $\mathcal{U}$ be a countable open cover of A. Then $\mathcal{U} = \{U_1, U_2, \cdots\}$ where each element $U_k \in \mathcal{U}$ is an open subset of A. Since A is a subspace of X, every open set U_k in A is of the form $G \cap A$ for some open set G in X. Therefore, there exists open sets $V(U_k)$ for each U_k such that $A \cap V(U_k) = U_k$.[5]

Define $\mathcal{V} = \{X - A, V(U_1), V(U_2), \cdots\}$. Clearly, $\mathcal{V}$ is a countable open cover[6] of X. We have X is countably compact, thus $\mathcal{V}$ has a finite subcover, say $\mathcal{V}'$. Without loss of generality assume[7] that $X - A \in \mathcal{V}'$. Suppose $X - A \notin \mathcal{V}'$, then we can define another finite subcover $\mathcal{V}' \cup \{X - A\}$. Thus $\mathcal{V}' = \{X - A,\ V(U_{n_1}),\ V(U_{n_2}), \cdots,\ V(U_{n_k})\}$.

Then the corresponding subcover $\mathcal{U}' = \{U_{n_1}, U_{n_2}, \cdots, U_{n_k}\}$ is a finite subcover of $\mathcal{U}$. Since countable open cover $\mathcal{U}$ and closed subset A are arbitrary, every closed subset of X with relative topology is countably compact. Therefore, countable compactness is weakly hereditary. □

Remark. *Proof depends on the following,*

1. *There is an extension map, $\psi : P(A) \to P(X)$ that preserve open sets (and closed sets). This ψ is an open map which not a true inverse of the restriction, $r : P(X) \to P(A)$, defined by $r(G) = G \cap A$ for every subset G of X.*

2. *Also we rely on the subset A being closed. Suppose X have many countable open covers, but X has only uncountable open covers corresponding to a particular countable open cover of A. In such a case, X being countably compact is insufficient for A to be countably compact.*

The behaviour of countinous functions

We will now study the nature of continuous functions defined on countably compact spaces. Suppose X, Y are topological space and function $f : X \to Y$ is continuous. If X is countably compact, then $f(X)$ is also countably compact.

[4] Suppose A is not a proper subset of X. Then $X = A$ and A is countably compact.
[5] Relative topology, $\mathcal{T}_{/A} = \{G \cap A : G \in \mathcal{T}\}$
[6] $X - A$ is open in X. If $y \notin A$, then $y \in X - A$. If $y \in A$, then $y \in U_k$ for some k.
[7] Otherwise, you will have to consider two cases: $X - A \in \mathcal{V}'$ and $X - A \notin \mathcal{V}'$

Continuous images of countably compact spaces are countably compact. In other words, countable compactness is preserved under continuous functions.[8]

Theorem 11.2. *Countable compactness is preserved under continuous functions.[Joshi, 1983, 11.1.2]*

Synopsis. *Let X be countably compact and $f : X \to Y$ be continuous. Suppose $\mathcal{U}$ is a countable cover of $f(X)$, then X has a countable cover $\mathcal{V}$ obtained by taking inverse images. Since X is countably compact, $\mathcal{V}$ has a finite subcover $\mathcal{V}'$. Now taking images of members of $\mathcal{V}'$, we get a finite subcover $\mathcal{U}'$ of $f(X)$.*

Proof. Suppose X is a countably compact space, Y is a topological space and $f : X \to Y$ is a continuous function. Let $\mathcal{U} = \{U_1, U_2, \cdots\}$ be a countable cover of $f(X)$ by set open in $f(X)$. We have to show that $\mathcal{U}$ has a finite subcover.

Define $\mathcal{V} = \{f^{-1}(U_1), f^{-1}(U_2), \cdots\}$. Then $\mathcal{V}$ is a countable open cover of X, since $f^{-1}(U_k)$ are open subsets of X and,

$$\begin{aligned} \bigcup_{k=1}^{\infty} U_k = f(X) &\implies f^{-1}\left(\bigcup_{k=1}^{\infty} U_k\right) = X \\ &\implies \bigcup_{k=1}^{\infty} f^{-1}(U_k) = X \end{aligned}$$

We have, $\mathcal{V}$ is a countable open cover of X, which is a countably compact space. Therefore $\mathcal{V}$ has a finite subcover $\mathcal{V}' = \{f^{-1}(U_{n_1}),\ f^{-1}(U_{n_2}), \cdots,\ f^{-1}(U_{n_k})\}$.

$$\begin{aligned} \bigcup_{j=1}^{k} f^{-1}(U_{n_j}) = X &\implies f^{-1}\left(\bigcup_{j=1}^{k} U_{n_j}\right) = X \\ &\implies \bigcup_{j=1}^{k} U_{n_j} = f(X) \end{aligned}$$

Clearly $\mathcal{U}' = \{U_{n_1}, U_{n_2}, \cdots, U_{n_k}\}$ is a finite subcover of $\mathcal{U}$. Thus every countable open cover of $f(X)$ by sets open in $f(X)$ has a finite subcover. Therefore, continuous images of countably compact spaces are countably compact. □

Remark. *1. For a continuous function, $f : X \to Y$ the inverse images of open sets are open in X. The relation $f^{-1} \subset f(X) \times X$ is not a function. However, we may consider a function, $\psi : P(Y) \to P(X)$ such that $\psi(U) = f^{-1}(U)$ for any subset U of Y. This ψ is an open map which maps open subsets of Y to open subsets of X.*

Theorem 11.3. *Every continuous, real-valued function on a countably compact, metric space is bounded and attains its extrema.[Joshi, 1983, 11.1.7]*

[8] A topological property is preserved under continuous functions if whenever a space has that property so does every continuous image of it.[Joshi, 1983, 6.1.9]

Synopsis. *Let X be a countably compact space and function $f : X \to \mathbb{R}$ be continuous. Then $f(X) \subset \mathbb{R}$ is countably compact. Real line $\mathbb{R}$ is metrisable*[9]*. Then $f(X)$ is countably compact, metric space. Therefore $f(X)$ compact.*[10]*. The subset $f(X)$ of $\mathbb{R}$ is bounded and closed, since every compact subset of $\mathbb{R}$ is bounded and closed. Thus $f(X)$ contains its supremum and infimum. Therefore, f is bounded and attains its extrema.*

Proof. Let X be a countably compact space and $f : X \to \mathbb{R}$ be continuous, real-valued function on the countably compact space, X. We have to show that f is bounded and attains its extrema.

Since countable compactness is preserved under continuous functions, $f(X)$ is countably compact subset of $\mathbb{R}$. Since, $f(X)$ is a subset of the metric space, $\mathbb{R}$ and metrisability is hereditary, $f(X)$ is again metrisable. (suppose) We have, every countably compact, metric space is compact. Then $f(X)$ is a compact subset of $\mathbb{R}$.

Since every compact subset of $\mathbb{R}$ is bounded and closed, $f(X)$ is bounded and closed. Since every closed subset of $\mathbb{R}$ contains supremem and infimum, $f(X)$ contains its extrema. Therefore, every continuous, real-valued function on a countably compact space is bounded and attains its extrema.

We have assumed that every countably compact, metric space is compact. This result will be proved in the last section of this chapter. □

Remark. *Since countably compact, metric spaces are compact. The above theorem can be used to prove that continuous, real-valued functions on a compact, metric space attains its extrema.*

Due to the Lebesgue covering lemma, next result is quite simple.*

Theorem 11.4. *Every continuous, real-valued function on a countably compact, metric space is uniformly continuous.*

Proposition 11.5. *Let X be a first countable, Hausdorff space. Then every countably compact subset A of X is closed.[Joshi, 1983, Exercises 11.1.7]*

11.1.2 Sequential Compactness

Weakly hereditary property

Theorem 11.6. *Sequential compactness is weakly hereditary.[Joshi, 1983, Exercises 11.1.3]*

The behaviour of countinous functions

Theorem 11.7. *Sequential compactness is preserved under continuous functions.[Joshi, 1983, Exercises 11.1.4]*

[9] [Joshi, 1983, 4.2 Example 4], $\mathbb{R}$ with usual metric $d : R \to R,\ d(x, y) = |x - y|$

[10] [Joshi, 1983, 11.1.11] On metric spaces, countable compactness $\implies$ compactness.

Synopsis. *Let X be sequentially compact and function $f : X \to Y$ be continuous. Then any sequence, $\{y_k\}$ in $f(X)$ has a sequence, $\{x_k\}$ in X such that $f(x_k) = y_k$. Sequence $\{x_k\}$ has a subsequence $\{x_{n_k}\}$ converging to x, then sequence $\{f(x_n)\}$ in $f(X)$ has the subsequence $\{f(x_{n_k})\}$ converging to $f(x)$.*

Proof. Let X be a sequentially compact space, function $f : X \to Y$ be continuous and $\{y_n\}$ be a sequence in $f(X)$ subset of Y. Construct a sequence $\{x_n\}$ such that $f(x_k) = y_k$, $\forall k$.

Every sequence in X has a convergent subsequence. Thus $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ converging to $x \in X$. The image of this subsequence $\{f(x_{n_k})\}$ is a subsequence of $\{y_k\}$. We claim that, $\{f(x_{n_k})\}$ converges to $f(x) \in f(X)$.

Let U be an open set containing $f(x)$, then $f^{-1}(U)$ is an open set containing x. Since $\{x_{n_k}\}$ converges to x. There exists an integer n such that for every $k \geq n$, $x_k \in f^{-1}(U)$. Clearly, for each $k \geq n$, $f(x_k) \in U$. Since U is arbitrary, $\{f(x_{n_k})\}$ converges to $f(x)$. Therefore, the image of any sequentially compact space is sequentially compact. In other words, sequentially compactness is preserved under continuous functions. □

Remark. *1. Given a sequence $\{y_n\}$ in $f(X)$, there is a sequence of subsets $\{U_n\}$ in $P(Y)$ such that $U_n = f^{-1}(y_n)$. Since each U_n is non-empty, we can construct a sequence $\{x_n\}$ in X using a choice function. The convergent subsequence of $\{y_n\}$ depends on the selection of this choice function.*

Given every sequentially compact, metric space is countably compact. We may assert the properties of countably compact, metric spaces on sequentially compact, metric spaces.

Theorem 11.8. *Every continuous, real-valued function on a sequentially compact, metric space is bounded and attains its extrema.*

Theorem 11.9. *Every continuous, real-valued function on a sequentially compact, metric space is uniformly continuous.[Joshi, 1983, Exercises 11.1.6]*

11.1.3 Countable Compactness on T_1 spaces

In this section, we are going to see four different characterisations of countable compactness in T_1 spaces. The first two characterisations doesn't have anything to do with the T_1 axiom.

T_1 Space A topological space X satisfy T_1 axiom if for any two distinct points $x, y \in X$, there exists an open set $U \subset X$ containing x but not y.[Joshi, 1983, 7.1.2]

countable compactness A topological space is countably compact if every countable open cover has a finite subcover.[Joshi, 1983, 11.1.1]

finite intersection property A family $\mathcal{F}$ of subsets of X has finite intersection property(f.i.p.) if every finite subfamily of $\mathcal{F}$ has a non-empty intersection.[Joshi, 1983, 10.2.6]

accumulation point A point $x \in X$ is accumulation point of a subset $A \subset X$ if every open set containing x has atleast one point of A other than x.[Joshi, 1983, 5.2.7]

limit point A point $x \in X$ is a limit point of a sequence $< x_k >$ in X if for every open set U containing x, there exists an integer $N \in \mathbb{N}$ such that $x_k \in U$ for every $k \geq N$.[Joshi, 1983, 4.1.7]

cluster point A point $x \in X$ is a cluster point of a sequence $< x_k >$ in X if for any neighbourhood V of x, the sequence $< x_k >$ assumes a point in V infinitely many times.[11]

Countable compactness in T_1 spaces

Theorem 11.10. *In a T_1 space X, following statements are equivalent,*

1. *X is countably compact*
2. *Every countably family of closed subsets of X with finite intersection property have non-empty intersection.*
3. *Every infinite subset $A \subset X$ has an accumulation point.*[12]
4. *Every sequence $< x_k >$ in X has a cluster point.*
5. *Every infinite open cover of X has a proper subcover.[Arens-Dugundji]*

Proof. $1 \implies 2$
Suppose X is countably compact. Let $\mathcal{C} = \{C_1, C_2, \cdots\}$ be a countable family of closed subsets of X with empty intersection. Define $\mathcal{U} = \{X - C_1, X - C_2, \cdots\}$ is a family of open subsets of X. By de Morgan's law, [13]

$$\bigcap_{k=1}^{\infty} C_k = \phi, \text{ then } X = X - \left(\bigcap_{k=1}^{\infty} C_k\right) = \bigcup_{k=1}^{\infty} (X - C_k)$$

We have $\mathcal{U}$ is a countable cover of X and X is countably compact space. Thus $\mathcal{U}$ has a finite subcover $\mathcal{U}' = \{X - C_{n_1}, X - C_{n_2}, \cdots, X - C_{n_k}\}$.

$$\mathcal{U}' \text{ is a cover of } X, \text{ then } X = \bigcup_{j=1}^{k} \left(X - C_{n_j}\right)$$

$$X - \bigcup_{j=1}^{k} \left(X - C_{n_j}\right) = \bigcap_{j=1}^{k} \left(X - \left(X - C_{n_j}\right)\right) = \bigcap_{j=1}^{k} C_{n_j} = \phi$$

Now $\mathcal{C}' = \{C_{n_1}, C_{n_2}, \cdots, C_{n_k}\}$ has empty intersection. This is a contradiction to the finite intersection property of $\mathcal{C}$. Thus $\mathcal{C}$ has non-empty intersection. Therefore, every countably family of closed subsets of X have non-empy intersection.

---

[11] x is a cluster point of $< x_k >$ if for every integer N, there exists $k > N$ such that $x_k \in V$.In other words, $< x_k >$ is frequently in V.[Joshi, 1983, 10.1.9]

[12] Every infinite subset of $\mathbb{R}$ has a limit point is equivalent to the completeness axiom.

[13] Complement of Intersection = Union of complements, $X - (C \cap D) = (X - C) \cup (X - D)$,

$2 \implies 1$

Let $\mathcal{U} = \{U_1, U_2, \cdots\}$ be a countable cover of X. Then $\mathcal{C} = \{X - U_1, X - U_2, \cdots\}$ is a countable family of closed subsets of X.

Let $\mathcal{U}' = \{U_{n_1}, U_{n_2}, \cdots, U_{n_k}\}$ be any finite subfamily of $\mathcal{U}$. Suppose X is not countably compact, then $\mathcal{U}$ doesn't have a finite subcover. Therefore, $\mathcal{U}'$ is not a cover of X. And $\mathcal{C}$ is a family of closed sets with finite intersection property.

Therefore by assumption, the countable family of closed sets $\mathcal{C}$ has a non-empty intersection.

$$\bigcap_{k=1}^{\infty} C_k \neq \phi, \text{ then } \bigcap_{k=1}^{\infty} C_k = \bigcap_{k=1}^{\infty} (X - U_k) = X - \left(\bigcup_{k=1}^{\infty} U_k \right) \neq \phi$$

Then $\mathcal{U}$ is not a cover of X as well. This is a contradiction, therefore X is countably compact.

$1 \implies 3$
Suppose X is countably compact. Let A be an infinite subset of X. Suppose A doesn't have an accumulation point.

Let B be a countably infinite subset of A. Then B also doesn't have any accumulation point. Therefore, the derived set B' is empty. Thus B is a closed subset of X. Since countable compactness is weakly hereditary, subspace B is again countably compact.

For each point $b \in B$, there is an open set V_b such that $V_b \cap B = \{b\}$, since $b \in B$ is not an accumulation point. Thus $\mathcal{U} = \{V_b \cap B : b \in B\}$ is a countable open cover of B. Clearly, $\mathcal{U}$ doesn't have any finite subcover.

This is a contradiction to B being countably compact. Therefore, A has an accumulation point. □

11.1.4 Variations of Compactness on Metric Spaces

In this document, we will see that from metric space point of view these two notions were equivalent to the compactness and were used alternatively. For example : in functional analysis (semester 3), you will find definitions like 'a normed space is compact iff every sequence in it has a convergent subsequence', which is clearly sequential compactness for a topologist.

Lindeloff A topological space is Lindeloff iff every open cover has a countable subcover.

First countable A topological space is first countable iff every point in it has a countable local base.

Second countable A topological space is second countable iff it has a countable base.

Base A family of subsets $\mathcal{B}$ of X is a base of a topological space if every open set can be expressed as union of some members of $\mathcal{B}$

Base Characterisation A family of subsets $\mathcal{B}$ of X is a base of a topological space iff for every $x \in X$, and for every neighbourhood U of x, there is a member $B \in \mathcal{B}$ such that $x \in B \subset U$.

Local Base A family of subsets $\mathcal{L}$ of X is a local base at point $x \in X$ if for every neighbourhood U of x, there is a member $L \in \mathcal{L}$ such that $x \in L \subset U$.

Equivalence

We are going to see when these three notions: compactness, countable compactness and sequentially compactness are equivalent.

Theorem 11.11. *Countably compact, metric spaces are second countable.*

Synopsis. *For every positive real number r, there exists a non-empty maximal subsets A_r with every pair of points atleast r distance apart. A_r are finite. The union of maximal subsets $A_{\frac{1}{n}}$ for each natural number n is a countable, dense subset D of X. Thus countably compact, metric spaces are separable. The family $\mathcal{B}$ of all open balls with center at $d \in D$ and rational radius is a countable, base for X. Thus countably compact, metric spaces are second countable.*

Proof. Let $(X; d)$ be a countably compact,, metric space. For each positive real number $r \in \mathbb{R},\ r > 0$ construct a family of subsets $A_r \subset X$ such that it is a maximal set of points which are atleast r distances apart.

Then A_r is finite for every positive real number r. Suppose A_r is infinite for some real number $r > 0$, then A_r has a accumulation point, say x by the Characterisation of countable compactness of X.

Then every neighbourhood of x must intersect A_r at infinitely many points, since every metric space is a T_1 space. Consider $B(x, \frac{r}{2})$. Since any two points of $B(x, \frac{r}{2})$ are less than r distances apart, the intersection $B(x, \frac{r}{2}) \cap A_r$ can have atmost one point in it. Thus for every positive real number r, A_r is finite.

Define $D = \cup_{n=1}^{\infty} A_{\frac{1}{n}}$. We claim that D is a countable, dense subset of X.

Let $x \in X$ and $B(x, r)$ be an open ball containing x, then there exists integer $n \in \mathbb{N}$ such that $\frac{1}{n} < r$.[14]

Then $B(x, r) \cap D \neq \phi$, since $B(x, r) \cap A_{\frac{1}{n}} \neq \phi$. Suppose $B(x, r) \cap A_{\frac{1}{n}} = \phi$, then $A_{\frac{1}{n}}$ is not maximal. Since, x is atleast $r > \frac{1}{n}$ distance apart from each points of $A_{\frac{1}{n}}$. Therefore, D intersects with every open set and thus dense in X.

We have have a countable, dense subset D of X. Therefore, X is separable. Now define $\mathcal{B} = \{B(x, r) : r \in \mathbb{Q},\ x \in D\}$. Clearly, $\mathcal{B}$ is a countable base for X. By the construction of $\mathcal{B}$, X is second countable.[15] □

[14] By archimedean property of integers, we have $\forall r \in \mathbb{R},\ r > 0,\ \exists n \in \mathbb{N}$ such that $nr > 1$.

[15] Every separable, metric space is second countable.

Countable Compactness, Lindeloff $\iff$ Compactness

Theorem 11.12. *A topological space X is compact iff it is countably compact, Lindeloff space.*

Proof. Let X be a compact space. Let $\mathcal{U}$ be a countable open cover of X, then $\mathcal{U}$ has a finite subcover $\mathcal{U}'$. Therefore, every compact space is countably compact.[16]

Conversely, suppose X is a countably compact, Lindeloff space. Since X is Lindeloff, every open cover $\mathcal{U}$ has a countable subcover $\mathcal{U}'$. Since X countably compact, every countable open cover $\mathcal{U}'$ has a finite subcover $\mathcal{U}''$. Thus every open cover $\mathcal{U}$ has a finite subcover $\mathcal{U}''$. Therefore every countably compact, Lindeloff space is compact. □

Countable Compactness, First Countable $\implies$ Seq. Compactness

Theorem 11.13. *Every countably compact, first countable space is Sequentially compact.*

Proof. Let X be a countably compact, first countable space. Let $\{x_n\}$ be a sequence in X. By, equivalent conditions[17] of countably compact spaces, every sequence in countably compact space X has a cluster point, say x. We have, X is first countable. Therefore, X has a countable local base $\mathcal{L}$ at $x \in X$. How to construct a subsequence of $\{x_n\}$ converging to x ?$\star$[18] □

Remark. *Every sequentially compact space is countably compact.*$\star$

Theorem 11.14. *In a second countable space, all the three forms of compactness are equivalent.[Joshi, 1983, 11.1.10]*

Proof. Every second countable space is both first countable and Lindeloff. Every countably compact, Lindeloff space is countably compact. Therefore every countably compact, second countable space compact. Again, every countably compact, first countable space is sequentially compact. Therefore every countably compact, second countable space is sequentially compact. Conversely, every compact space is countably compact and every sequentially compact space is countably compact.[19] □

Theorem 11.15. *In a metric space, all the three forms of compactness are equivalent.[Joshi, 1983, 11.1.11]*

Proof. In a metric space each form of compactness implies second countability. And in second countable spaces, they are all equivalent. □

[16] Countable compactness is a weaker notion than compactness.

[17] [Joshi, 1983, 11.1] Conditions 1,2, and 4 are equivalent. 2 $\implies$ 4 without T_1 axiom is out of scope.

[18] [Joshi, 1983, Exercises 10.1.11]

[19] Countable compactness is a weaker notion than sequential compactness as well.

Chapter 12

The Fundamental Group

12.1 Homotopy of Paths

Definitions 12.1. *Let X, Y be topological spaces and $f : X \to Y$, $f' : X \to Y$ be continuous functions. Then f, f' are homotopic if there exists a continuous function $F : X \times I \to Y$ such that for every $x \in X$, $F(x,0) = f(x)$ and $F(x,1) = f'(x)$. And we write, $f \simeq f'$.*

Definitions 12.2. *Let X be a topological space and $f : I \to X$ and $f' : I \to X$ be two paths. Then f, f' are path-homotopic if they have same initial point x_0 (ie, $x_0 = f(0) = f'(0)$), same final point x_1 (ie, $x_1 = f(1) = f'(1)$) and they are homotopic (ie, $\exists F : I \times I \to X$ such that $\forall x \in I$, $F(x,0) = f(x)$ and $F(x,1) = f'(x)$ also fixed at the end points x_0 and x_1(ie, $\forall t \in I$, $F(0,t) = x_0$ and $F(1,t) = x_1$). And we write $f \simeq_p f'$.*

Remark. *If two paths f, $f'$are homotopic, then they have the same end points and there exists a (topologically) continuous deformation from one path into another.*

Proposition 12.3. *The relations $\simeq$, $\simeq_p$ are equivalence relations.*

Proof. Homotopy : Let f, f' be continuous functions from X into Y. Then f and f' are homotopic, $f \simeq f' \iff \exists F : X \times I \to Y$ such that F is continuous, $F(x,0) = f(x)$, and $F(x,1) = f'(x)$

1. $f \simeq f$
   We have $f : X \to Y$ is continuous. Define $F : X \times I \to Y$ such that $F(x,t) = f(x)$. Clearly, F is continuous, $F(x,0) = f(x)$ and $F(x,1) = f(x)$. And $\exists F : X \times I \to Y \implies f \simeq f$.

2. $f \simeq f' \implies f' \simeq f$
   We have, $f \simeq f'$. Thus there exists a continuous function $F : X \times I \to Y$ such that $F(x,0) = f(x)$ and $F(x,1) = f'(x)$.
   Consider $F' : X \times I \to Y$ defined by $F'(x,t) = F(x,1-t)$. Clearly, F' is continuous, $F'(x,0) = F(x,1) = f'(x)$, and $F'(x,1) = F(x,0) = f(x)$. Thus, $\exists F'(x,t) : X \times I \to Y \implies f' \simeq_p f$

3. $f \simeq f',\ f' \simeq f'' \implies f \simeq f''$
   We have, $f \simeq f' \iff \exists F : X \times I \to Y$ such that F is continuous,

$F(x,0) = f(x)$ and $F(x,1) = f'(x)$.

Similarly, $f' \simeq f'' \iff \exists F' : X \times I \to Y$ such that F' is continuous, $F'(x,0) = f'(x)$ and $F'(x,1) = f''(x)$.

Consider $G : X \times I \to Y$ defined by

$$G(x,t) = \begin{cases} F(x,2t) & ,t \in [0,\frac{1}{2}] \\ F'(x,2t-1) & ,t \in [\frac{1}{2},1] \end{cases}$$

We have, $G(x,\frac{1}{2}) = F(x,1) = F'(x,0) = f'(x)$. Thus, G is continuous by pasting lemma since $[0,\frac{1}{2}] \cap [\frac{1}{2},1] = \{\frac{1}{2}\}$. Also $G(x,0) = F(x,0) = f(x)$ and $G(x,1) = F'(x,1) = f''(x)$. Thus, $\exists G : X \times I \to Y \implies f \simeq f''$

Path Homotopy : Let $f,\ f',\ f''$ be paths in X. Then f and f' are path homotopic, $f \simeq_p f' \iff \exists F : I \times I \to X$ such that F is continuous, $\forall s \in [0,1],\ F(s,0) = f(s),\ F(s,1) = f'(s)$ and $\forall t \in [0,1],\ F(0,t) = f(0) = f'(0),\ F(1,t) = f(1) = f'(1)$

1. $f \simeq_p f$
   We have $f : I \to X$ is continuous. Define $F : I \times I \to X$ such that $\forall s,t \in [0,1],\ F(s,t) = f(s)$. Clearly, F is continuous, $\forall s \in [0,1],\ F(s,0) = f(s),\ F(s,1) = f(s)$ and $\forall t \in [0,1],\ F(0,t) = f(0),\ F(1,t) = f(1)$. Thus, $\exists F : I \times I \to X \implies f \simeq_p f$.

2. $f \simeq_p f' \implies f' \simeq_p f$
   We have, $f \simeq_p f'$. Thus there exists a continuous function $F : I \times I \to X$ such that $\forall s \in [0,1],\ F(s,0) = f(s),\ F(s,1) = f'(s)$ and $\forall t \in [0,1],\ F(0,t) = f(0) = f'(0),\ F(1,t) = f(1) = f'(1)$.

   Consider $F' : I \times I \to X$ defined by $F'(s,t) = F(s,1-t)$. Clearly, F' is continuous. And $F'(s,0) = F(s,1) = f'(s)$, and $F'(s,1) = F(s,0) = f(s)$. Also, $F'(0,t) = F(0,1-t) = f(0) = f'(0)$ and $F'(1,t) = F(1,1-t) = f(1) = f'(1)$. Thus, $\exists F'(s,t) : I \times I \to X \implies f' \simeq_p f$

3. $f \simeq f',\ f' \simeq f'' \implies f \simeq f''$
   We have, $f \simeq f' \iff \exists F : I \times I \to X$ such that F is continuous, $\forall s \in [0,1],\ F(s,0) = f(s),\ F(s,1) = f'(s)$ and $\forall t \in [0,1],\ F(0,t) = f(0) = f'(0),\ F(1,t) = f(1) = f'(1)$

   Similarly, $f' \simeq f'' \iff \exists F' : I \times I \to X$ such that F' is continuous, $\forall s \in [0,1],\ F'(s,0) = f'(s),\ F'(s,1) = f''(s)$ and $\forall t \in [0,1],\ F'(0,t) = f'(0) = f''(0),\ F'(1,t) = f'(1) = f''(1)$

   Consider $G : I \times I \to X$ defined by

$$G(s,t) = \begin{cases} F(s,2t) & ,t \in [0,\frac{1}{2}] \\ F'(s,2t-1) & ,t \in [\frac{1}{2},1] \end{cases}$$

We have, $G(s, \frac{1}{2}) = F(s, 1) = F'(s, 0) = f'(s)$. Thus, G is continuous by pasting lemma[Munkres, 2003, §18.3 pp. 106], since $[0, \frac{1}{2}] \cap [\frac{1}{2}, 1] = \{\frac{1}{2}\}$.

Also $G(s, 0) = F(s, 0) = f(s)$ and $G(s, 1) = F'(s, 1) = f''(s)$.

Again, $\forall t \in [0, \frac{1}{2}],\ G(0, t) = F(0, 2t) = f(0) = f'(0) = f''(0)$ and $\forall t \in [\frac{1}{2}, 1],\ G(0, t) = F'(0, 2t - 1) = f(0) = f'(0) = f''(0)$. Therefore, $\forall t \in [0, 1],\ G(0, t) = f(0) = f''(0)$.

Similarly, $\forall t \in [0, \frac{1}{2}],\ G(1, t) = F(1, 2t) = f(1) = f'(1) = f''(1)$ and $\forall t \in [\frac{1}{2}, 1],\ G(1, t) = F'(1, 2t - 1) = f(1) = f'(1) = f''(1)$. Therefore, $\forall t \in [0, 1],\ G(1, t) = f(1) = f''(1)$. Thus, $\exists G : I \times I \to X \implies f \simeq_p f''$

□

Definitions 12.4. *Let f be a path in X (ie, $f : I \to X$), then $[f]$ is the equivalence class of all paths homotopic to f in X. (ie, $g \in [f] \iff f \simeq_p g$)*

Remark. *The set of homotopy classes of functions from X into Y is dentoed by $[X, Y]$. And, the set of all path-homotopic classes on X is denoted by $[I, X]$.*

Remark (Straight-line homotopy)**.** *[Munkres, 2003, §51 Example 1 pp. 320] Let X be a topological space, and f, g be continuous functions from X into a eucilidean space, say $\mathbb{R}^2$. Then f, g are straight line homotopic if there exists a continuous function F from $X \times I$ such that F deforms f into g along straight line segments joining them.*

For example, $F(x, t) = (1 - t)f(x) + tg(x)$.

Remark. *Let A be a convex subspace of $\mathbb{R}^n$. Then any two paths in A from x_0 to x_1 are path homotopic in A.*

Proof. —continue page 321— □

Remark. *[Munkres, 2003, §51 Example 2 pp. 321] This demonstrates that the straight-line homotopy is very sensitive to the holes in the space.*

Definitions 12.5. *Let f, g be two paths in X (ie, $f : I \to X$ and $g : I \to X$) such that $f(0) = x_0$, $f(1) = g(0) = x_1$ and $g(1) = x_2$. Then the prduct $h = f * g$ is given by $h : I \to X$ and*

$$h(s) = \begin{cases} f(2s) & , s \in [0, \frac{1}{2}] \\ g(2s - 1) & , s \in [\frac{1}{2}, 1] \end{cases}$$

This h is well-defined, and continuous by pasting lemma.[1]

Definitions 12.6. *The product operation on path-homotopy classes is defined by $[f] * [g] = [f * g]$.*

[1]Pasting Lemma : Let $X = A \cup B$, where A and B are closed in A. Let $f : A \to Y$ and $g : B \to Y$ be continuous. If $f(x) = g(x)$ for every $x \in A \cup B$, then f and g combine to give a continuous function $h : X \to Y$, defined by setting $h(x) = f(x)$ if $x \in A$ and $h(x) = g(x)$ if $x \in B$.

Remark. *The product of path-homotopic classes is well-defined.*

Proof. Let F be a path-homotopy between $f,\ f' \in [f]$ and G be a path-homotopy between $g,\ g' \in [g]$. Then $H : I \times I \to X$ defined by

$$H(s,t) = \begin{cases} F(2s,t) & s \in [0, \frac{1}{2}] \\ G(2s-1,t) & s \in [\frac{1}{2}, 1] \end{cases}$$

Then H is well-defined, and continuous by pasting lemma.

$\forall s \in [0, \frac{1}{2}],\ H(s,0) = F(2s,0) = f(2s)$ and
$\forall s \in [\frac{1}{2}, 1],\ H(s,0) = G(2s-1,0) = g(2s-1)$.
$\implies H(s,0) = (f * g)(s)$, by the definition of $f * g$

$\forall s \in [0, \frac{1}{2}],\ H(s,1) = F(2s,1) = f'(2s)$ and
$\forall s \in [\frac{1}{2}, 1],\ H(s,1) = G(2s-1,1) = g'(2s-1)$.
$\implies H(s,1) = (f' * g')(s)$, by the definition of $f' * g'$

$H(0,t) = F(0,t) = f(0) = x_0 = (f * g)(0)$, and
$H(1,t) = G(1,t) = g'(1) = x_2 = (f' * g')(1)$

Then $H : I \times I \to X$ is a path-homotopy between $f * g$ and $f' * g'$. □

Definitions 12.7 (Groupoid)**.** *Let G be a set and $*$ be a binary operation on G. Then $(G, *)$ is a groupoid if it satisfies the following axioms*

g1 Associativity - $\forall x, y, z \in G,\ (x * y) * z = x * (y * z)$

g2 Existence of left and right identies - There exist unique elements e_L and e_R such that $\forall x \in G,\ x * e_R = x$ and $e_L * x = x$.

g3 Existence of inverse
$\forall x \in G,\ \exists x^{-1} \in G$ such that $x * x^{-1} = e_L$ and $x^{-1} * x = e_R$

Definitions 12.8 (Positive Linear Map)**.** *A positive liear map $p : [a,b] \to [c,d]$ is the unique map of the form $p(x) = mx + k$ such that $p(a) = c$ and $p(b) = d$. Clearly, scaling factor, $m = \frac{d-c}{b-a}$ as we want to transform an interval of length $b-a$ into an interval of length $d-c$. And offset k is given by,*

$$p(a) = \frac{d-c}{b-a}a + k = c \implies k = c - \frac{a(d-c)}{b-a} = \frac{bc-ad}{b-a}$$

But, we won't fix m and k in $p(x) = mx+k$, instead we will focus on the unique map with graph of positive slope and passing through required end points. The graph of a positive linear map from $[a,b]$ to $[c,d]$ is always a straight-line with positive slope.

Remark. *The inverse of a positive linear map is also a positive linear map. Given $p : [a,b] \to [c,d], p(x) = mx + k$, where $m = \frac{d-c}{b-a},\ k = \frac{bc-ad}{b-a}$. Then it's inverse, $\bar{p} : [c,d] \to [a,b]$ is given by $p(y) : m'y + k'$, where $m' = \frac{b-a}{d-c} = \frac{1}{m},\ k' = \frac{ad-bc}{d-c} = \frac{-k}{m}$. Clearly $m > 0 \implies m' = \frac{1}{m} > 0$.*

Remark. *The composite of two positive linear maps is also a (piece-wise) positive linear map. Let f, g be two positive linear maps. Then their composite map $f * g$ is given by*

$$(f*g)(x) = \begin{cases} f(2x) & x \in [a, \frac{a+b}{2}] \\ g(2\left(x - \frac{b-a}{2}\right)) & x \in [\frac{a+b}{2}, b] \end{cases}$$

Remember $f * g$ exists only if $f(b) = g(a)$. Therefore, $f * g$ is a well-defined, continuous (by pasting lemma) and (piecewise) positive linear map.

Lemma 12.9. *Let f, f' be two paths in X and $k : X \to Y$ be a continuous function. Let F be the path homotopy in X between the paths f and f'. Then $k \circ F$ is a path homotopy in Y between that paths $k \circ f$ and $k \circ f'$ That is, path homotopy is preserved under a continuous function.*

Lemma 12.10. *Let f, g be two paths in X with $f(1) = g(0)$ and $k : X \to Y$ be a continous function. Then $k \circ (f * g) = (k * f) \circ (k * g)$*

Theorem 12.11. *Let f, g, h be paths in a topological space X, and $[f], [g], [h]$ be respective path-homotopy classes. Suppose the operation product, $*$ is defined by*

$$[f] * [g] = [f *' g] \text{ where } (f *' g)(s) = \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ g(2s-1) & s \in [\frac{1}{2}, 1] \end{cases}$$

Then the product, $*$ has the following properties :

1. *Associativity*

$$\forall [f], [g], [h] \in [I, X],\ ([f] * [g]) * [h] = [f] * ([g] * [h])$$

2. *Existence of left and right identies*
   Let $e_x : I \to X$ defined by $\forall s \in [0, 1],\ e_x(s) = x$. Let f be a path from x_0 to x_1, then there exist unique paths e_{x_0} and e_{x_1} such that $[f] * [e_{x_1}] = [f]$ and $[e_{x_0}] * [f] = [f]$. That is, e_{x_0}, e_{x_1} are respectively the left and right path-homotopy-identities.

3. *Existence of inverse*
   Let f be a path in X. The path,$\bar{f}$, defined by $\bar{f}(s) = f(1-s)$ is the reverse path of f. Then $[f] * [\bar{f}] = [e_{x_0}]$ and $[\bar{f}] * [f] = [e_{x_1}]$. That is, the inverse of class of f is the class of reverse path of f.

In other words, Set of all path-homotopy classes together with binary operation product, $*$ is a groupoid. ie, $([I, X], *)$ is a groupoid.

Proof. Step 1 : Properties 2&3

Let $e_0 : I \to I$ such that $e_0(t) = 0,\ \forall t \in I$. And $i : I \to I$ such that $i(t) = t,\ \forall t \in I$. Then $e_0 * i$ is also a path. Since I is convex, there is a path homotopy[2] G between i and $e_0 * i$. Let $f : I \to X$ be continuous path in X from

[2] $G : I \times I \to I,\ G(s, 0) = i(s),\ G(s, 1) = (e_0 * i)(s),\ G(0, t) = 0,\ G(1, t) = 1.$

x_0 to x_1. Then $f \circ G$ is a path homotopy (by Lemma 2) in X between $f \circ i$ and $f \circ e_0 * i$ where $f \circ i$ and $f \circ e_0$ are paths from x_0 to x_1 in X.

$$\begin{aligned} f \circ (e_0 * i) =& (f \circ e_0) * (f \circ i), \text{ by Lemma 1} \\ =& e_{x_0} * f, \text{ since } \forall s \in I,\ f(e_0(s)) = x_0 = e_{x_0}(s) \text{ and } f * i \simeq_p f \end{aligned}$$

Therefore $[e_{x_0}] * [f] \simeq_p [f]$, since $e_0 * i \simeq_p i$, and $f \circ (e_0 * i) \simeq_p f \circ i = f$.

Similarly, $e_1 : I \to I$ such that $e_1(t) = 1$. Let H be a path homotopy[3] between $i * e_1$ and i. Thus, $f \circ H$ is a path homotopy in X from $f \circ (i * e_1)$ and $f \circ i$.

$$\begin{aligned} f \circ (i * e_1) =& (f \circ i) * (f \circ e_1), \text{ by Lemma 1} \\ =& f * e_{x_1}, \text{ since } f * i \simeq_p f,\ i * e_1 \simeq_p e_1,\ \forall s \in I,\ (f(e_1(s)) = x_1 = e_{x_1}(s) \end{aligned}$$

Since $i * e_1 \simeq_p i$, we have $f \circ (i * e_1) \simeq_p f \circ i = f$. Therefore $[f] * [e_{x_1}] \simeq_p [f]$. Thus, $[f] * [e_{x_1}] \simeq_p [f] \simeq_p [e_{x_0}] * [f]$. Therefore, $[f]$ has left and right inverses ie, property 2 holds.

Consider inverse path $\bar{i} : I \to I,\ \bar{i}(s) = 1 - s$. Then $i * \bar{i}$ is a path in I with both end points at 0. We have, $e_0 : I \to I,\ e_0(s) = 0$ is also a path with both end points at 0. Since I is convex, there is a path homotopy H in I between e_0 and $i * \bar{i}$. Then $f \circ H$ is a path homotopy between $f \circ e_0 = e_{x_0}$ and $f \circ (i * \bar{i}) = (f \circ i) * (f \circ \bar{i}) = f * \bar{f}$. Therefore, $[e_{x_0}] \simeq_p [f] * [\bar{f}]$.

Similarly $\bar{i} * i$ and e_1 are paths with both end points at 1. Since I is convex, there is a path homotopy G in I between $\bar{i} * i$ and e_1. Then $f \circ G$ is a path homotopy between $f \circ (\bar{i} * i) = (f \circ \bar{i}) * (f \circ i) = \bar{f} * f$ and $f \circ e_1 = e_{x_1}$. Therefore, $[\bar{f}] * [f] \simeq_p [e_{x_1}]$. Thus the path $\bar{f} : I \to X,\ \bar{f}(s) = f(1 - s),\ \forall s \in I$ is reverse of f. Also $[f] * [\bar{f}] = [e_{x_0}]$ and $[\bar{f}] * [f] = [e_{x_1}]$. ie, property 3 holds.

Step 2 : Property 1

Let f, g, h be three paths in X and $f(1) = g(0) = x_1$ and $g(1) = h(0) = x_2$. Then $f * (g * h)$ is defined by

$$(g * h)(s) = \begin{cases} g(2s) & s \in [0, \frac{1}{2}] \\ h(2s-1) & s \in [\frac{1}{2}, 1] \end{cases}$$

$$\begin{aligned} (f * (g * h))(s) &= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ (g * h)(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\ &= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ g(2(2s-1)) & s \in [\frac{1}{2}, \frac{3}{4}] \\ h(2(2s-1)-1) & s \in [\frac{3}{4}, 1] \end{cases} \end{aligned}$$

Similarly, $(f * g) * h$ is defined by,

$$(f * g)(s) = \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ g(2s-1) & s \in [\frac{1}{2}, 1] \end{cases}$$

[3] $H : I \times I \to I,\ H(s, 0) = (i * e_1)(s),\ H(s, 1) = i(s),\ H(0, t) = 0,\ H(1, t) = 1$

$$((f*g)*h)(s) = \begin{cases} (f*g)(2s) & s \in [0, \frac{1}{2}] \\ h(2s-1) & s \in [\frac{1}{2}, 1] \end{cases}$$
$$= \begin{cases} f(2(2s)) & s \in [0, \frac{1}{4} \\ g(2(2s-1)) & s \in [\frac{1}{4}, \frac{1}{2}] \\ h(2s-1) & s \in [\frac{1}{2}, \frac{3}{4}] \end{cases}$$

Clearly, $f*(g*h)$ and $(f*g)*h$ are distinct path with common endpoints. ie, $(f*(g*h))(0) = f(0) = ((f*g)*h)(0)$. And $(f*(g*h)(1) = h(1) = ((f*g)*h)(1)$.

We need to define a path homotopy G between $f*(g*h)$ and $(f*g)*h$. Let $[a,b],[c,d] \subset I$. Consider the path $p : I \to I$ defined by the following three unique[4] positive linear maps $p_1 : [0,a] \to [0,c]$, $p_2 : [a,b] \to [c,d]$ and $p_3 : [b,1] \to [d,1]$.

$$p(t) = \begin{cases} p_1(t) & t \in [0,a] \\ p_2(t) & t \in [a,b] \\ p_3(t) & t \in [b,1] \end{cases}$$

We can easily construct, a path homotopy P between identity map $i : I \to I$, $i(s) = s$ and p as follows :

$$P(s,t) = \begin{cases} t + (p_1(t) - t)\frac{s}{a} & s \in [0,a] \\ t + (p_2(t) - t)\frac{(s-a)}{(b-a)} & s \in [a,b] \\ t + (p_3(t) - t)\frac{(s-b)}{(1-b)} & s \in [b,1] \end{cases}$$

Therefore, we have $f*(g*h) \simeq_p i$ since there exists a path homotopy P corresponding to $[a,b] = [\frac{1}{2}, \frac{3}{4}]$ and $[c,d] = [x_1, x_2]$. Similarly $(f*g)*h \simeq_p i$ since there exists a path homotopy P where $[a,b] = [\frac{1}{4}, \frac{1}{2}]$ and $[c,d] = [x_1, x_2]$. ie, $[f*(g*h)] \simeq_p [(f*g)*h]$. □

Theorem 12.12. *Let f be a path in X, and $a_0, a_1, \cdots, a_n$ be numbers such that $0 = a_0 < a_1 < \cdots < a_n = 1$. Let $f_i : I \to X$ be the path that equals the positive linear map of I onto $[a_{i-1}, a_i]$ followed by f. Then $[f] = [f_1] * [f_2] * \cdots [f_n]$. In other words, every path is path-homotopic to a piecewise-linear path.*

Proof. Let f be a piece-wise positive linear map such that

$$f(t) = \begin{cases} f_1(t) & t \in [0 = a_0, a_1] \\ f_2(t) & t \in [a_1, a_2] \\ \vdots & \vdots \\ f_n(t) & t \in [a_{n-1}, a_n] \end{cases}$$

where $f_i : I \to [a_{i-1}, a_i]$ such that $f_i(t)$ are a positive linear maps.

Consider the path $p : I \to I$ defined by the unique positive linear maps on the subintervals of any partition $\{0 = x_0, x_1, \cdots, x_n\}$ of I.

[4] $p_1(t) = \frac{ct}{a}$, $p_2(t) = \frac{(d-c)t}{b-a} + \frac{bc-ad}{b-a}$, $p_3(t) = \frac{(1-d)t}{1-b} + \frac{d-b}{1-b}$

ie, $0 = x_0 < x_1 < \cdots < x_n = 1$

$$\begin{aligned} p_1 : [x_0, x_1] &\to [a_0, a_1] \\ p_2 : [x_1, x_2] &\to [a_1, a_2] \\ &\vdots \\ p_n : [x_{n-1}, x_n] &\to [a_{n-1}, a_n] \end{aligned}$$

$$\text{Define, } p(t) = \begin{cases} p_1(t) & t \in [x_0, x_1] \\ p_2(t) & t \in [x_1, x_2] \\ \vdots & \vdots \\ p_n(t) & t \in [x_{n-1}, x_n] \end{cases}$$

Then there exists a path homotopy P between identity map $i : I \to I,\ i(t) = t$ and p given by

$$P(s,t) = \begin{cases} t + (p_1(t) - t)\frac{a_1}{x_1} & s \in [0, x_1] \\ t + (p_2(t) - t)\frac{s - x_1}{x_2 - x_1} & s \in [x_1, x_2] \\ \vdots & \\ t + (p_n(t) - t)\frac{s - x_{n-1}}{x_n - x_{n-1}} & s \in [x_{n-1}, x_n] \end{cases}$$

Since any product of $f_1,\ f_2,\ \cdots,\ f_n$ is a path p for some partition decided by the order of associativity. This partition can be constructed as follows :
Let the last product operation (by associtivity) corresponds to $\frac{1}{2}$. The expression on its left corresponds to $[0, \frac{1}{2}]$ and expression on the right corresponds to $[\frac{1}{2}, 1]$. If there are any operations on any of these parts, the last operation (by associtivity) in them corresponds to the midpoint the respective subinterval and so on.

For examples : Consider, $(f_1 * (f_2 * f_3)) * (f_4 * f_5)$. Suppose we number the operations, $(f_1 *_1 (f_2 *_2 f_3)) *_3 (f_4 *_4 f_5)$. Then we have, $*_3 \to \frac{1}{2} \implies *_1 \to \frac{1}{4} \implies *_2 \to \frac{3}{8}$. Again $*_3 \to \frac{1}{2} \implies *_4 \to \frac{3}{4}$. Thus, we have $\{0, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{3}{4}, 1\}$.

Thus given two paths f and f' with distinct order of associativity of n paths : $f_1,\ f_2,\ \cdots,\ f_n$. We have path homotopy P, P' given by the $P(s,t)$ for the respective partition constructed according to the order of associativity. Then, we have $f \simeq_p p$ and $f' \simeq_p p$. Thus, irrespective of the order of associativity all these paths are path homotopic. ie, $[f] = [f_1] * [f_2] \cdots [f_n]$ □

Part II

ME010203 Numerical Analysis with Python

Chapter 13

Module III

Definitions 13.1. *Given* $(n+1)$ *data points* $(x_k, y_k),\ k = 0, 1, \cdots, n,$ *the problem of estimating* $y(x)$ *using a function* $y : \mathbb{R} \to \mathbb{R}$ *that satisfy the data points is the interpolation problem. ie,* $y(x_k) = y_k,\ k = 0, 1, \cdots, n.$

Definitions 13.2. *Given* $(n+1)$ *data points* $(x_k, y_k),\ k = 0, 1, \cdots, n,$ *the problem of estimating* $y(x)$ *using a function* $y : \mathbb{R} \to \mathbb{R}$ *that is sufficiently close to the data points is the curve-fitting problem.*
ie, Given $\epsilon > 0,\ |y(x_k) - y_k| < \epsilon,\ k = 0, 1, \cdots, n.$

Remark. *The data could be from scientific experiments or computations on mathematical models. The interpolation problem assumes that the data is accurate. But, curve-fitting problem assumes that there are some errors involved which are sufficiently small.*

Definitions 13.3. *Given* $(n+1)$ *data points* $(x_k, y_k),\ k = 0, 1, \cdots, n,$ *the problem of estimating* $y(x)$ *using a polynomial function of degree* n *that satisfy the data points is the polynomial interpolation problem.*

Remark. *Polynomial is the 'simplest' interpolant.[Kiusalaas, 2013, 3.2]*

13.1 Polynomial Interpolation

There exists a unique polynomial of degree n that satisfy $(n+1)$ distinct data points. There are a few methods to find this polynomial : 1. Lagrange's method 2. Newton's method.

13.1.1 Lagrange's Method

Interpolation polynomial[1] is given by,

$$P(x) = \sum_{i=0}^{n} y_i l_i(x), \text{ where } l_i(x) = \prod_{j=0, j\neq i}^{n} \frac{x - x_i}{x_j - x_i} \tag{13.1}$$

[1] Using P_n to represent some polynomial of degree n. It is quite a confusing a notation when it comes to Newton's method as author construct a psuedo-recursive definition.

Remark. *Lagrange's cardinal functions l_i, are polynomials of degree n and*

$$l_i(x_j) = \delta_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}$$

Proposition 13.4. *Error in polynomial interpolation is given by*

$$f(x) - P(x) = \frac{(x - x_0)(x - x_1)\cdots(x - x_n)}{(n+1)!} f^{(n+1)}(\xi) \tag{13.2}$$

where $\xi \in (x_0, x_n)$

Remark. *The error increases as x moves away from the unknown value ξ.*

13.1.2 Newton's Method

The interpolation polynomial is given by,

$$P(x) = a_0 + a_1(x - x_0) + \cdots + a_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}) \tag{13.3}$$

where $a_i = \nabla^i y_i,\ i = 0, 1, \cdots, n$.

Remark. *For Newton's Method, usually it is assumed that $x_0 < x_1 < \cdots < x_n$.*

Remark. *Lagrange's method is conceptually simple. But, Newton's method is computationaly more efficient than Lagrange's method.*

Computing coefficients a_i of the polynomial

The coefficients are given by,

$$a_0 = y_0,\ a_1 = \nabla y_1,\ a_2 = \nabla^2 y_2,\ a_3 = \nabla^3 y_3, \cdots, a_n = \nabla^n y_n \tag{13.4}$$

Remark. *The divided difference $\nabla^i y_i$ are computed as follows:*

$$\nabla y_1 = \frac{y_1 - y_0}{x_1 - x_0}$$

$$\nabla y_2 = \frac{y_2 - y_1}{x_2 - x_1} \qquad \nabla^2 y_2 = \frac{\nabla y_2 - \nabla y_1}{x_2 - x_1}$$

$$\nabla y_3 = \frac{y_3 - y_2}{x_3 - x_2} \qquad \nabla^2 y_3 = \frac{\nabla y_3 - \nabla y_2}{x_3 - x_2} \qquad \nabla^3 y_3 = \frac{\nabla^2 y_3 - \nabla^2 y_2}{x_3 - x_2}$$

x_0	y_0				
x_1	y_1	∇y_1			
x_2	y_2	∇y_2	$\nabla^2 y_2$		
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\ddots$	
x_n	y_n	∇y_n	$\nabla^2 y_n$	$\cdots$	$\nabla^n y_n$

Table 13.1: The $\nabla^i y_i$ Computation Table

Remark. *Practise Problems*
Find interpolation polynomial for the following data points :

1. $\{(0,7),\ (2,11),\ (3,28)\}$
   Ans : $5x^2 - 8x + 7$
   [Kiusalaas, 2013, Example 3.1]
2. $\{(-2,-1),\ (1,2),\ (4,59),\ (-1,4),\ (3,24),\ (-4,-53)\}$
   Ans : $x^3 - 2x + 3$
   [Kiusalaas, 2013, Example 3.2]
3. $\{(-1.2,-5.76),\ (0.3,-5.61),\ (1.1,-3.69)\}$
   Ans : $x^2 + x - 6$
   [Kiusalaas, 2013, Problem Set 3.1.1]
4. $\{(-3,0),\ (2,5),\ (-1,-4),\ (3,12),\ (1,0)\}$
   Ans : $x^2 + 2x - 3$
   [Kiusalaas, 2013, Problem Set 3.1.7]
5. $\{(0,1.225),(3,0.905),(6,0.652)\}$
   Ans : $0.0037x^2 - 0.1178x + 1.225$
   [Kiusalaas, 2013, Problem Set 3.1.9]

Remark. *In Lagrange's Method, we can interpolate at the given point even without computing the polynomial. In Newton's method, we have to compute polynomial and then interpolate for the given point.*

That is, evaluate the value of cardinal polynomials at the point and substitute in Equation 13.1 as shown in Section 3.2.[Kiusalaas, 2013, Example 3.1]

13.1.3 Implementation of Newton's Method

Program 13.5. *Computing Coefficients*

```
def coefficients(xData,yData):
  m = len(xData)
  a = yData.copy()
  for k in range(1,m):
    a[k:m] = (a[k:m]-a[k-1])/(xData[k:m]-xData[k-1])
  return a
```

Line 1 `def coefficients(xData,yData):`
Defines a function which takes two arguments/parameters, named $xData$ and $yData$. In [Kiusalaas, 2013, 3.2], you will find coeffts which I have changed to coefficients. $xData$,$yData$ are numpy array objects. $xData$ is a array with values $x_0, x_1, \cdots, x_n$. And $yData$ is array with values $y_0, y_1, \cdots, y_n$. For example, the value of x_3 can be accessed as $xData[3]$.

Line 2 `m = len(xData)`
The functon $len()$ is extended by numpy to give the length of array objects. In this context, $len(xData)$ will return the value $n+1$, since there are $n+1$ values in $xData$ array.

Line 3 `a = yData.copy()`
We need a copy of $yData$ to work with. Unlike other programming languages like java, in python $a = yData$ will assign a new label a to the same memory location and manipulating a will corrupt the original data in $yData$ as well. In order to avoid this, we are **making a copy of the array object using the array method provided by the numpy library.**

Line 4 `for k in range(1,m):`
This is a python loop statement. This ask python interpreter to repeat the following sub-block $m-1$ times.[2] In this context, Line 5 will be executed n times, since the $range(1,m)$ object is a list-type object with values $1, 2, \cdots, m-1$. And intepreter executes Line 5 for each values in the $range()$ object, ie, $k = 1, 2, \cdots, m-1$ before interpreting Line 6.

Line 5 `a[k:m] = (a[k:m]-a[k-1])/(xData[k:m]-xData[k-1])`
This is very nice feature available in python. **This statement, evaluates $m-k$ values in a single step.ie, $a[k], a[k+1], \cdots, a[m]$. This calculation corresponds to subsequent columns of the divided difference table, that we are familiar with.** For example, executing Line 5 with $k = 3$ is same as evaluating the $\nabla^3 y_j$ column. Note that the value $a[0]$ is never updated and similarly $a[2]$ changes when Line 5 is executed with $k = 1, 2$. From column 3 onward, $a[2]$ is not updated. Therefore, **after completing nth executing of the Line 5, we have $a[0] = y_0,\ a[1] = \nabla y_1,\ a[2] = \nabla^2 y_2, \cdots,\ a[n] = \nabla^n y_n$.**

Line 6 `return a`
This returns the array a which is the array of coefficients.

The logic of this program is in Line 4 and Line 5. So they need more explanation/understanding than anything else.

Program 13.6. *Interpolating using Newton's Method*

```
def interpolate(a,xData,x):
  n = len(xData)-1
  p = a[n]
  for k in range(1,n+1):
    p = a[n-k]+(x-xData[n-k])*p
  return p
```

The logic this program is in Line 3, Line 4 and Line 5.

Line 3 : We initialize the polynomial with the coefficient $a[n] = \nabla^n y_n = a_n$.

Line 4 : We are going define the polynomial recursively. This takes exactly n steps further. So we use a loop which repeats n times.

Line 5 : The value of p and k changes each time Line 5 is executed. Let P_j be the value in p after executing Line 5 with $k = j$. Then,

[2] Python block is a group of statement with same level of indentation. A sub-block is a block with an additional indentation.

$P_0 = p = a[n]$
$P_1 = a[n-1] + (x - x_{n-1})P_0$
$P_2 = a[n-2] + (x - x_{n-2})P_1$
$\vdots$
$P_n = a[0] + (x - x_0)P_{n-1}$.
Clearly, P_n is the unique n degree polynomial given by the Newton's method.

Program 13.7. *How to interpolate ?*

```
from numpy import array
xData = array([-2,1,4,-1,3,-4])
yData = array([-1,2,59,4,24,-53])
a = coefficients(xData,yData)
print(interpolate(a,xData,2))
```

You will have to define both the functions (coefficients, interpolate) before doing this.

Line 1 `from numpy import array`
For defining array objects, we need to import them from numpy library.

Line 2 `xData = array([-2,1,4,-1,3,-4])`
You can change this line according to the first component of the given data points.

Line 3 `yData = array([-1,2,59,4,24,-53])`
You can change this line according to the second component of the given data points.

Line 4 `a = coefficients(xData,yData)`
Call function `coefficients` and store the array returned into a

Line 5 `print(interpolate(a,xData,2))`
Call function `interpolate` to interpolate at $x = 2$ and print the value $P(2)$

Program 13.8 (Just for Fun)**.** *We can do more using sympy !*

```
from numpy import array}
from sympy import Symbol
xData = array([-2,1,4,-1,3,-4])
yData = array([-1,2,59,4,24,-53])
a = coefficients(xData,yData)
x = Symbol('x')
p = interpolate(a,xData,x)
p.subs({x:2})
```

Remark. *Programming Problems*

1. $\{(0.15, 4.79867),\ (2.30, 4.49013),\ (3.15, 4.2243),\ (4, 85, 3.47313),\ (6.25, 2.66674),\ (7.95, 1.51909)\}$ *[Kiusalaas, 2013, Example 3.4]*
2. $\{(0, -0.7854),\ (0.5, 0.6529),\ (1, 1.7390),\ (1.5, 2.2071),\ (2, 1.9425)\}$ *[Kiusalaas, 2013, Problem Set 3.1.5]*

13.1.4 Limitations of Polynomial Interpolation

1. Inaccuracy - The error in interpolation increases as the point moves away from most of the data points.

2. Oscilation - As the number of data points considered for polynomial interpolation increases, the degree of the polynomial increases. And the graph of the interpolant tend to oscilate excessively. In such cases, the error in interpolation is quite high.

3. The best practice is to consider four to six data points nearest to the point of interest and ignore the rest of them.

Remark. *The interpolant obtained by joining cubic polynomials corresponding to four nearest data points each, is a cubic spline*[3].

13.2 Roots of a Function

Definitions 13.9. *Let $f : \mathbb{R} \to \mathbb{R}$, then $x \in \mathbb{R}$ is a root of f if $f(x) = 0$.*

Remark. *Suppose $a < b$ and $f(a), f(b)$ are nonzero and are of different signs. If f is continuous in $[a, b]$, then there is a point $c \in [a, b]$ such that $f(c) = 0$.*

Thus given $a < b$ and $f(a), f(b)$ are nonzero values of different sign, then there may be a bracketed root in $[a, b]$.

Note : There is no guarantee that there exists a root in $[a, b]$ as we are not sure about the continuity of f.

Remark. *Given a bracketed root, we can find it using*

1. *Bisection Method or*

2. *Newton-Raphson Method*

13.2.1 Bisection Method

Suppose $a < b$ and $f(a), f(b)$ are nonzero values of different signs. We evaluate $f(c)$ where $c = \frac{a+b}{2}$. If $f(c)$ is a nonzero value, then at least one of the pairs $f(a), f(c)$ or $f(c), f(b)$ are of different signs. WLOG suppose that $f(a), f(c)$ are of different signs, then set $b = c$ and $c = \frac{a+b}{2}$. And continue this process until we get sufficiently accurate value of a root.

Remark. *Suppose $f(x) = x^5 - 2$. Then $f(0) = -2,\ f(1) = -1,\ f(2) = 30$. Since f is known to be continuous, there is a bracketed root in $[1, 2]$. Now*
$f(1.5) > 0 \implies [1, 1.5]$
$f(1.25) > 0 \implies [1, 1.25]$
$f(1.125) < 0 \implies [1.125, 1.25]$
$f(1.1825) > 0 \implies [1.125, 1.1825]$
$f(1.15375) > 0 \implies [1.125, 1.15375]$
$f(1.139375) < 0 \implies [1.139375, 1.15375]$

[3] Cubic spline is a function, the graph of which is piece-wise cubic

$f(1.1465625) < 0 \implies [1.1465625, 1.15375]$
$f(1.150156250) > 0 \implies [1.1465625, 1.15015625]$
$f(1.148359375) < 0 \implies [1.1483594, 1.15015625]$
$f(1.149257825) > 0 \implies [1.1483594, 1.14925783]$

Thus, we have 1.14 *is a root of* f *with accuracy upto two decimal points.*

13.2.2 Newton-Raphson Method

Suppose f is differentiable at $x \in \mathbb{R}$ and $f(x) \neq 0$. Then compute $x = x - \frac{f(x)}{df(x)}$ and evaluate $f(x)$. Repeat this process to get more accurate value of a root near x.

Remark. *Suppose* $f(x) = x^5 - 2$. *Then* $df(x) = 5x^4$. *Let* $x = 2$. *Then*

$$\begin{aligned}
x = 2 - \frac{30}{80} &\implies f(1.625) = 9.330 \\
x = 1.625 - \frac{9.330}{34.86} &\implies f(1.35735) = 2.6074 \\
x = 1.35735 - \frac{2.6074}{16.9721} &\implies f(1.20373) = 0.52733 \\
x = 1.20373 - \frac{0.52733}{10.4975} &\implies f(1.15351) = 0.04224 \\
x = 1.15351 - \frac{0.042245}{8.85225} &\implies f(1.148738) = 0.00034312
\end{aligned}$$

Thus we have 1.1487 *is quite close to a root of* f.

Chapter 14

Module IV

Consider a system of n linear, simultaneous equations in n unknowns,

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= b_2 \\ &\vdots \\ A_{n1}x_1 + A_{n2}x_2 + \cdots + A_{nn}x_n &= b_n \end{aligned}$$

We may represent them using matrices as $Ax = b$. That is,

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Gauss Elimination and Doolittle Decomposition are two methods to solve sytem of equaitons, $Ax = b$.

14.1 Gauss elimination method

Gauss elimination method has of two phases 1. elimination and 2. back substitution. In elimination phase, system $Ax = b$ is transformed into an equivalent system $Ux = c$ where U is an upper-triangular[1] matrix. And in back substitution phase, $Ux = c$ is solved. Since $Ax = b$ and $Ux = c$ are equivalent, they have the same solution x.

14.1.1 Elimination Phase

We can eliminate unknowns from an equation by adding a scalar multiple of an equation to another equation of the system. In matrices, this is equivalent to adding a scalar multiple of one row to another row, say $R_i \leftarrow R_i + \lambda R_k$.

$$\begin{array}{r} A_{k1}x_1 + A_{k2}x_2 + \cdots + A_{kn}x_n = b_k + \\ \lambda\left(A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n = b_i\right) \\ \hline (A_{k1} + \lambda A_{i1})x_1 + (A_{k2} + \lambda A_{i2})x_2 + \cdots + (A_{kn} + \lambda A_{in})x_n = b_k + \lambda b_i \end{array}$$

[1] upper triangular - all the entries below the main diagonal are zero. ie $U_{ij} = 0$, if $i < j$

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kn} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i1} & A_{i2} & \cdots & A_{in} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \xrightarrow{R_i \leftarrow R_i + \lambda R_k} \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{kn} \\ \vdots & \vdots & \ddots & \vdots \\ A_{i1} + \lambda A_{k1} & A_{i2} + \lambda A_{k2} & \cdots & A_{in} + \lambda A_{kn} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$\begin{bmatrix} \vdots \\ b_k \\ \vdots \\ b_i \\ \vdots \end{bmatrix} \xrightarrow{R_i \leftarrow R_i + \lambda R_k} \begin{bmatrix} \vdots \\ b_k \\ \vdots \\ b_i + \lambda b_k \\ \vdots \end{bmatrix}$$

14.1.2 Back substitution

Let $Ux = c$ be a system of n linear equations in n unknowns and U is an upper triangular matrix. Then we can solve the system of equations from the back.

$$\begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,n-1} & U_{1,n} \\ 0 & U_{2,2} & \cdots & U_{2,n-1} & U_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & U_{n-1,n-1} & U_{n-1,n} \\ 0 & 0 & \cdots & 0 & U_{n,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix}$$

$$U_{n,n}\ x_n = c_n \implies x_n = \frac{c_n}{U_{n,n}}$$

$$\sum_{i=n-1}^{n} U_{n-1,i}\ x_i = c_{n-1} \implies x_{n-1} = \frac{c_{n-1} - U_{n-1,n} x_n}{U_{n-1,n-1}}$$

$$\cdots$$

$$\sum_{i=1}^{n} U_{1,i}\ x_i = c_1 \implies x_1 = \frac{c_1 - \sum_{i=2}^{n} U_{1,i} x_i}{U_{1,1}}$$

14.1.3 Illustrative example

Consider the following system of linear equations,

$$\begin{aligned} 4x_1 - 2x_2 + x_3 &= 11 \\ -2x_1 + 4x_2 - 2x_3 &= -16 \\ x_1 - 2x_2 + 4x_3 &= 17 \end{aligned}$$

We may represent the above system of linear equations using matrices,

$$\begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 11 \\ -16 \\ 17 \end{bmatrix}$$

Phase 1 : Elimination Process

Using eq.1, the unknown x_1 is eliminated from all subsequent equations. An equivalent operation can be performed on both the matrices A and b by adding a suitable scalar multiples of row R_1 to row R_2 and R_3.

$$\begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix} \xrightarrow[R_3 \leftarrow R_3 - 0.25R_1]{R_2 \leftarrow R_2 + 0.5R_1} \begin{bmatrix} 4 & -2 & 1 \\ 0 & 3 & -1.5 \\ 0 & -1.5 & 3.75 \end{bmatrix}$$

$$\begin{bmatrix} 11 \\ -16 \\ 17 \end{bmatrix} \xrightarrow[R_3 \leftarrow R_3 - 0.25R_1]{R_2 \leftarrow R_2 + 0.5R_1} \begin{bmatrix} 11 \\ -10.5 \\ 14.25 \end{bmatrix}$$

And using eq.2, x_2 is eliminated from all subsequent equations(only those rows below it). Again, we perform this by adding suitable scalar multiples of row 2 to row R_3.

$$\begin{bmatrix} 4 & -2 & 1 \\ 0 & 3 & -1.5 \\ 0 & -1.5 & 3.75 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 + 0.5R_2} \begin{bmatrix} 4 & -2 & 1 \\ 0 & 3 & -1.5 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\begin{bmatrix} 11 \\ -16 \\ 17 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 + 0.5R_2} \begin{bmatrix} 11 \\ -10.5 \\ 9 \end{bmatrix}$$

The elimination process is complete when all entries below the diagonal elements are reduced to zero. ie, upper triangular matrix.

Phase 2 : Back substitution Process

The unknowns are easily found from the equations by solving them in the reverse order. The unknowns are solved from the bottom and solved variables are used to solve the remain unknowns.

$$\begin{bmatrix} 4 & -2 & 1 & 11 \\ 0 & 3 & -1.5 & -10.5 \\ 0 & 0 & 3 & 9 \end{bmatrix} \rightarrow \begin{cases} 4x_1 - 2x_2 + & x_3 = 11 \\ 3x_2 - & 1.5x_3 = -10.5 \\ & 3x_3 = 9 \end{cases}$$

$$\begin{aligned} x_3 &= \frac{9}{3} = 3 \\ x_2 &= \frac{-10.5 + 1.5x_3}{3} = -2 \\ x_1 &= \frac{11 - x_3 + 2x_2}{4} = 1 \end{aligned}$$

Remark. *Why don't they use row-reduced echelon matrix of A to simplify the back substitution phase ?*

This doesn't have much advantage from algorithmic point of view. That is, the time complexity (number of steps for computation) is unaffected. And algorithms always prefer methods even with slight advantage in time or memory.

And they won't consider complications in the manual execution of the method. Therefore, programmers won't consider alternate algorithm for the sake of computational simplicity.

14.1.4 Python : Gauss elimination method

Program 14.1 (Gauss elimination).

```
from numpy import dot
def gaussElimination(a,b):
  n = len(b)
  for k in range(0,n-1):
    for i in range(k+1,n):
      if a[i,k] != 0.0:
        lam = a[i,k]/a[k,k]
        a[i,k+1:n] = a[i,k+1:n]-lam*a[k,k+1:n]
        b[i] = b[i]-lam*b[k]
  for k in range(n-1,-1,-1):
    x[k] = (b[k]-dot(a[k,k+1:n],x[k+1:n]))/a[k,k]
  return b
```

Line 1 `from numpy import dot`
Imports the "dot()" function for numpy arrays which takes two 'numpy arrays' as input arguments and returns the dot product of them.

Line 2 `def gaussElimination(a,b):`
it defines "gaussElimination()" as a function which takes two arguments (inputs). First argument is the coefficient matrix A and second argument is the constant matrix b of the linear system of the form $Ax = b$.

Line 3 $n = len(b)$
it assigns the length of the list b into variable n which is obviously the number of equations.

Line 4 `for` k `in range(0,`$n-1$`):`
it is a loop construct. Five instructions following it are part of this loop body, which are executed for each values of k ie, $k = 0, 1, \cdots, n-1$. For each value of k, the unknown x_{k+1} is selected for elimination process.

Line 5 `for i in range(`$k+1$`,`n`):`
it is a loop inside another loop. Four instructions following it are part of this loop body, which are executed for each values of i, ie, $i = k+1, k+2, \cdots, n$. This eliminates x_{k+1} from all the equations after the $k+1$th equation of the system. Value of $i+1$ is the equation[2] from which x_{k+1} is eliminated.

Line 6 `if` $a[i,k] != 0.0$`:`
If $a[i,k] = A_{i+1,k+1} \neq 0$, then those three instruction following it are executed. Otherwise, it skips the execution of those three statements. If $x_{k+1} = x[k]$ is not there in the ith equation, it doesn't need to be eliminated.

[2]Python starts counting from zero. For example : $A_{11} = a[0,0]$, $x_1 = x[0]$ and $b_1 = b[0]$

Line 7 $lam = a[i,k]/a[k,k]$
In this step, λ is computed so that equ.$(i+1)$ - λ equ.$(k+1)$ doesn't have x_{k+1} term in it.

Line 8 $a[i,k+1:n] = a[i,k+1:n] - lam \times a[k,k+1:n]$
Coefficients of $(i+1)$th equation are updated.
Equivalent to $a[i,0:n] = a[i,0:n] - lam \times a[k,0:n]$, since zeroes need not be substracted. This is same as equ.$(i+1) \leftarrow$ equ.$(i+1)$ - λ equ.$(k+1)$

Line 9 $b[i] = b[i] - lam * b[k]$
The same row operations are performed on the matrix b instead of using an augmented matrix.

Line 10 `for` k `in range(`$n-1,-1,-1$`):`
This is another loop construct. The following statement is executed n times for values of $k = n-1, n-2, \cdots, 0$. Value of $k+1$ gives the unknown x_{k+1} which is solved by the back substitution process.

Line 11 $x[k] = (b[k] - dot(a[k,k+1:n], x[k+1:n]))/a[k,k]$
This is the back substitution process. After elimination phase we have k equation in the form $A_{k,k}x_k + A_{k,k+1}x_{k+1} + \cdots + A_{k,n}x_n = b_k$. And we already have values of $x_{k+1},\ x_{k+2},\ \cdots,\ x_n$. Then

$$x_k = \frac{b_k - (A_{k,k+1}x_{k+1} + A_{k,k+2}x_{k+2} + \cdots)}{A_{k,k}}$$

This is equivalent to

$$b_k \leftarrow \frac{b_k - \begin{bmatrix} A_{k,k+1} & A_{k,k+2} & \cdots A_{k,n} \end{bmatrix} \begin{bmatrix} x_{k+1} \\ x_{k+2} \\ \vdots \\ x_n \end{bmatrix}}{A_{k,k}}$$

Remember : The values of x_k are updated into b_k as they are computed. Thus $x_k, x_{k+1}, \cdots, x_n$ are stored in b for next back substitution ie, for evaluating x_{k-1}. We start with x_{n-1}, as $x_n = b_n$ is already solved.

Line 12 `return b`
It returns the new b matrix as output of the "gaussElimination()" function where $x_k = b_k,\ \forall k$.

14.2 LU Decomposition Method : Doolittle

Let $Ax = b$ be a linear system of n equations in n unknowns and let $A = LU$ for some lower triangular matrix L and upper triangular matrix U. Then we have $LUx = Ly = b$ where $y = Ux$. There are two phases for this method : 1. LU decomposition and 2. substitution.

First, we compute L and U such that $A = LU$ using Gauss elimination. Then We can solve $Ly = b$ using forward substitution process and then solve

$Ux = y$ using back substitution process.

For Doolittle decomposition, we prefer to write A as a product LU as shown below:

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ L_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{n,1} & L_{n,2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,n} \\ 0 & U_{2,2} & \cdots & U_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & U_{n,n} \end{bmatrix}$$

$$A = \begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,n} \\ L_{2,1}U_{1,1} & L_{2,1}U_{1,2} + U_{2,2} & \cdots & L_{2,1}U_{1,n} + U_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ L_{n,1}U_{1,1} & L_{n,1}U_{1,2} + L_{n,2}U_{2,2} & \cdots & \sum_{k=1}^{n-1} L_{n,k}U_{k,n} + U_{n,n} \end{bmatrix}$$

Note that in Doolittle's decomposition method, the diagonal entries of the lower triangular matrix L are all 1. ie, $L_{ii} = 1$, $\forall i$. Thus, we can use an $n \times n$ matrix to represent both L and U by overwriting trivial entries(zeroes and ones) of both the matrices. And this matrix is reprensented by $[L\backslash U]$.[3]

$$[L\backslash U] = \begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,n} \\ L_{2,1} & U_{2,2} & \cdots & U_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ L_{n,1} & L_{n,2} & \cdots & U_{n,n} \end{bmatrix}$$

is the combined matrix made from both the triangular matrices L and U.

The triangular matrices L and U such that $LU = A$ can be computed the variables in the Gauss elimination method.

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ L_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{n,1} & L_{n,2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,n} \\ 0 & U_{2,2} & \cdots & U_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & U_{n,n} \end{bmatrix}$$

We can break down this matrix multiplication into the following row operations on the rows of the upper triangular matrix[4]

$$\begin{aligned} U_{R1} &\leftarrow U_{R1} \\ U_{R2} &\leftarrow L_{2,1} \cdot U_{R1} + U_{R2} \\ &\cdots \\ U_{Rn} &\leftarrow L_{n,1} \cdot U_{R1} + L_{n,2} \cdot U_{R2} + \cdots + L_{n,n-1} \cdot U_{R(n-1)} + U_{Rn} \end{aligned}$$

Clearly, λ we use to eliminate x_k from row i are $L_{i,k}$. And the matrix obtained after Gauss elimination is the upper triangular matrix U.

[3] algorithmic implementation all decomposition algorithms prefer to use a combined matrix

[4] U_{Rk} : kth row of the matrix U

14.2.1 Illustrative example

$$\text{Solve} \begin{bmatrix} -3 & 6 & -4 \\ 9 & -8 & 24 \\ -12 & 24 & -26 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 65 \\ -42 \end{bmatrix}$$

Phase 1 : LU Decomposition

Suppose, we have a system of three linear equations, then

$$A = LU = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix} \begin{bmatrix} U_{1,1} & U_{1,2} & U_{1,3} \\ 0 & U_{2,2} & U_{2,3} \\ 0 & 0 & U_{3,3} \end{bmatrix}$$

$$A = \begin{bmatrix} U_{1,1} & U_{1,2} & U_{1,3} \\ L_{2,1}U_{1,1} & L_{2,1}U_{1,2} + U_{2,2} & L_{2,1}U_{1,3} + U_{2,3} \\ L_{3,1}U_{1,1} & L_{3,1}U_{1,2} + L_{3,2}U_{2,2} & L_{3,1}U_{1,3} + L_{3,1}U_{2,3} + U_{3,3} \end{bmatrix}$$

We can compute L and U using the Gauss elimination process[5]. The matrix obtained after Gauss elimination on A is U and the values of the variable lam used in Gauss elimination are the entries in L. That is, in order to eliminate x_k from row i, we use $lam = L_{i,k}$.

$$\text{Given, } A = \begin{bmatrix} -3 & 6 & -4 \\ 9 & -8 & 24 \\ -12 & 24 & -26 \end{bmatrix}$$

$$\begin{bmatrix} -3 & 6 & -4 \\ 9 & -8 & 24 \\ -12 & 24 & -26 \end{bmatrix} \xrightarrow[R_3 \leftarrow R_3 - 4R_1]{R_2 \leftarrow R_2 + 3R_1} \begin{bmatrix} -3 & 6 & -4 \\ 0 & 10 & 12 \\ 0 & 0 & -10 \end{bmatrix} \implies L_{2,1} = 3,\ L_{3,1} = -4$$

We store these non-trivial entries of L into A itself.
That is, $A_{2,1} = L_{2,1},\ A_{3,1} = L_{3,1}$.

"In this case, $A_{3,2}$ became zero (this is not a trivial zero yet), and we won't eliminate x_2 from row 3 to save computation time. Thus, we are not computing $L_{3,2} = 0$ or storing it. However, the variable representing $L_{3,2}$ is $A_{3,2}$, which is already zero after Gauss elimination and we are quite happy with that."

Clearly, $L_{3,2} = 0$. Therefore, we have

$$U = \begin{bmatrix} -3 & 6 & -4 \\ 0 & 10 & 12 \\ 0 & 0 & -10 \end{bmatrix},\ L = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$$

Since we are already stored those two non-trivial entries of L into A. We get,

$$[L \backslash U] = \begin{bmatrix} -3 & 6 & -4 \\ 3 & 10 & 12 \\ -4 & 0 & -10 \end{bmatrix}$$

[5] We usually need a proof for such a strong statement. In this paper, they are more focussed on the application side and therefore we will don't present any vigorous proof.

Phase 2 : Substitution

Suppose $Ly = b$,

$$\begin{bmatrix} 1 & 0 & 0 \\ L_{2,1} & 1 & 0 \\ L_{3,1} & L_{3,2} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \rightarrow \begin{cases} y_1 & = b_1 \\ L_{2,1}y_1 + y_2 & = b_2 \\ L_{3,1}y_1 + L_{3,2}y_2 + y_3 & = b_3 \end{cases}$$

Now, we can find the values of y_k and store them into the matrix b itself.

$$b_1 \leftarrow y_1 = b_1$$
$$b_2 \leftarrow y_2 = b_2 - \begin{bmatrix} L_{2,1} \end{bmatrix} \begin{bmatrix} b_1 \end{bmatrix}, \text{ since } b_1 = y_1$$
$$b_3 \leftarrow y_3 = b_3 - \begin{bmatrix} L_{3,1} & L_{3,2} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \text{ since } b_1 = y_1,\ b_2 = y_2$$

In general,

$$b_k \leftarrow y_k = b_k - \begin{bmatrix} L_{k,1} & L_{k,2} & \cdots & L_{k,k-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k - 1 \end{bmatrix}, \text{ since } b_j = y_j,\ j = 1, 2, \cdots, (k-1)$$

We have $A = LU \implies LUx = b$. Suppose $Ux = y$, then we get $Ly = b$. First of all, we will solve $Ly = b$ using forward substitution.

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 65 \\ -42 \end{bmatrix} \rightarrow \begin{cases} y_1 & = -3 \\ 3y_1 + y_2 & = 65 \\ -4y_1 + y_3 & = -42 \end{cases}$$

$$y_1 = -3$$
$$y_2 = 65 - 3y_1 = 74$$
$$y_3 = -42 + 4y_1 = 54$$

Suppose $Ux = y$,

$$\begin{bmatrix} U_{1,1} & U_{1,2} & U_{1,3} \\ 0 & U_{2,2} & U_{2,3} \\ 0 & 0 & U_{3,3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Now, we can find the values of x_k and store them into the matrix y itself.

$$y_3 \leftarrow x_3 = \frac{y_3}{U_{3,3}}$$
$$y_2 \leftarrow x_2 = \frac{y_2 - \begin{bmatrix} y_3 \end{bmatrix} \begin{bmatrix} U_{2,3} \end{bmatrix}}{U_{2,2}}, \text{ since } y_3 = x_3$$
$$y_1 \leftarrow x_1 = \frac{y_1 - \begin{bmatrix} y_2 \\ y_3 \end{bmatrix} \begin{bmatrix} U_{1,2} & U_{1,3} \end{bmatrix}}{U_{1,1}}, \text{ since } y_2 = x_2,\ y_3 = x_3$$

In general,

$$x_k = \frac{y_k - \begin{bmatrix} y_{k+1} \\ y_{k+2} \\ \vdots \\ y_n \end{bmatrix} \begin{bmatrix} U_{k,k+1} & U_{k,k+2} & \cdots & U_{k,n} \end{bmatrix}}{U_{k,k}}, \text{ since } y_j = x_j,\ j = k{+}1, k{+}2, \cdots, n$$

14.2.2 Python : Doolittle's LU Decomposition method

Program 14.2.

```
1  from numpy import dot
2  def LUdecomposition(a):
3    n = len(a)
4    for k in range(0,n-1) :
5      for i in range(k+1,n):
6        if a[i,k] != 0.0:
7          lam = a[i,k]/a[k,k]
8          a[i,k+1:n] = a[i,k+1:n] - lam*a[k,k+1:n]
9          a[i,k] = lam
10   return a
11 def LUsolve(a,b):
12   n = len(a)
13   for k in range(1,n):
14     b[k] = b[k]-dot(a[k,0:k],b[0:k])
15   b[n-1] = b[n-1]/a[n-1,n-1]
16   for k in range(n-2,-1,-1):
17     b[k] = (b[k] - dot(a[k,k+1:n],b[k+1:n]))/a[k,k]
18   return b
```

This program mainly uses the Gauss elimination algorithm. Thus, the explanation for Lines 3-8 are not repeated here.

But remember the loop at Line 4 has inner loop at Line 5 and Line 7-9 are at same level of indentation which means they all are either executed or skipped depending on the truthness of the condition in Line 6. And Line 6-9 are executed for each instance of inner loop. Again, Line 5-9 are executed for each instance of the outer loop.

This time the gaussElimination() function which you have seen earlier is split into two functions 1. LUdecomposition() and 2. LUsolve(). And forward substituion is also added to LUsolve().

Line 2 `def LUdecomposition(a):`
LUdecomposition(A) computes L and U such that $A = LU$ and combine both triangular matrices into a single matrix $[L/U]$, by over-writting their trivial entries. And returns this combined matrix.

Line 9 $a[i,k] = lam$
Clearly, lam used for eliminating x_k from row i, $\lambda_{i,k} = a[i,k]/a[k,k] =$

$L[i,k],\ \forall k, \forall i,\ (i > k)$. Also $a[i,k]$ which is reduced zero by Gauss elimination process is not used anymore[6] in Gauss elimination process. Thus $L[i,k]$ can stored at $a[i,k]$ straight away. And $U[i,j],\ j \leq i$ are already the entries of the matrix obained from Gauss elimination. Thus for each iterations of k, the matrix a is updated ($k+1$th row and $k+1$th column) with respective entries of the combined matrix $[L\backslash U]$.

Line 10 `return` a
Matrix a is already $[L\backslash U]$, and thus LUdecomposition(A) returns $[L\backslash U]$ such that $A = LU$.

Line 11 `def LUsolve(a,b):`
LUsolve() function does both forward substitution and back substitution. Suppose $Ax = b$ is the system to be solved. Then the inputs of LUsolve() are $a = [L\backslash U]$ where $A = LU$.

Line 12 $n = len(a)$
We have to compute this again as this function starts fresh and thus value of the variable n from LUdecomposition() is lost.

Line 13 `for` k `in range(1,`n`):`
This is the loop for forward substitution.

Line 14 $b[k] = b[k] - dot(a[k, 0:k], b[0:k])$
updating b_{k^*} with y_{k^*} such that $Ly = b$ where $k^* = k - 1$.

$$b_k \leftarrow b_k - \begin{bmatrix} L_{k,1} & L_{k,2} & \cdots & L_{k,k-1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k - 1 \end{bmatrix}$$

Line 15 $b[n-1] = b[n-1]/a[n-1, n-1]$
Computing[7] y_n and storing it into b_n.

$$b_n \leftarrow \frac{b_n}{U_{n,n}}$$

Line 16 `for` k `in range(`$n-2$`,`-1`,`-1`):`
This is the loop for back substitution.

Line 17 $b[k] = (b[k] - dot(a[k, k+1:n], b[k+1:n]))/a[k,k]$
updating b_{k^*} with x_{k^*} such that $Ux = y$ where $k^* = k - 1$.

$$b_k \leftarrow \frac{b_k - \begin{bmatrix} x_{k+1} \\ x_{k+2} \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} U_{k,k+1} & U_{k,k+2} & \cdots & U_{k,n} \end{bmatrix}}{U_{k,k}}$$

[6] $A[i,k]$ is not used after elimination of x_k from row i - It turns out that the trivial zeroes which are ignored on the row operations in Gauss elimination not only save time, but also provide a variable to store our intermediate result $L_{i,k}$ in Doolittle method.

[7] Mathematically, you can define dot product of empty matrices as zero, but nympy dot function can't handle such a situation. Therefore, we have to do this step separately.

Line 18 `return` b
LUsolve($[L\backslash U]$,b) returns b where $b[i+1] = x_i$.

Programmer's Tip : There are few things to remember when spliting a function into two functions.

1. These functions are completely independent of one another.
2. Variables defined inside a function are not available outside.
3. The best way to give/take data to/from a function is through arguements/return-value

Beginner's Tip : In any programming language, we reuse variable. Thus, same variable may represent different values at different points of time. In Doolittle LU Decomposition, the variable 'a' initially represent matrix A, this variable is passed into LUdecomposition() function. In that function, A is changed to $[L\backslash U]$ in a step-by-step fashion. The value of 'a' in updated in step 7, 8 and 9. This is bit hard to imagine this transition of 'a' from A to $[L\backslash U]$ for a beginner at programming. Simiarly, in LUsolve() function, the variable 'b' changes from matrix b to matrix y, and then to matrix x.

14.3 Numerical Integraion

Numerical integration/Quadrature is the numerical approximation of $\int_a^b f(x)dx$ by $\sum_{i=0}^{n} A_i f(x_i)$ where x_i are nodal abscissas, and A_i are weights. There are two methods to determine these nodal abscissas and suitable weights so that the sum is sufficiently accurate to the value of the integral.

1. Newton-Cotes forumulas
2. Gauss quadrature

Newton-Cotes formulas are useful when $f(x)$ can be evaluated without much computation. And using those values $f(x)$ can be interpolated to a piecewise-polynomial function. Then using equally spaces nodal abscissas and suitable weights $\int_a^b f(x)dx$ can be numerically approximated.

Gauss quadrature rules require lesser evalutions of f. And therefore are quite useful when evaluation of $f(x)$ has much computational complexity. Also, this method can manage integrable singularities where as Newton-Cote formulas can't numerically integrate function with singularities.

14.3.1 Newton-Cotes formulas

We divide the interval of integral (a, b) into n subintervals of equal length, ie, $h = (b-a)/n$. Let $x_0 = a, x_1, x_2, \cdots, x_{n-1}, x_n = b$ be the end points of these subintervals. Then we can find an n degree polymonial interpolant satisfying f at those points, using Lagrange's method.

$$\text{Polynomial, } P(x) = \sum_{i=0}^{n} f(x_i) l_i(x) \text{ where } l_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_i}{x_i - x_j}$$

Thus the integral $I = \int_a^b f(x)dx$ can be numerically evaluated as follows:

$$\begin{aligned} I = \int_a^b P_n(x)dx &= \sum_{i=0}^{n} \left(f(x_i) \int_a^b l_i(x)dx \right) \\ &= \sum_{i=0}^{n} A_i f(x_i), \text{ where } A_i = \int_a^b l_i(x)dx \end{aligned}$$

The simplest cases of Newton-Cotes formulas are when $n = 1, 2, and 3$

Trapezoidal rule $n = 1 \implies A_0 = \frac{h}{2},\ A_1 = \frac{h}{2}$ and

$$\int_a^b f(x)dx = A_0 f(x_0) + A_1 f(x_1) = \frac{h}{2}(f(a) + f(b))$$

Simpson's $1/3$ **rule** $n = 2 \implies A_0 = \frac{h}{3},\ A_1 = \frac{4h}{3},\ A_2 = \frac{h}{3}$ and

$$\int_a^b f(x)dx = \sum_{i=0}^{2} A_i f(x_i) = \frac{h}{3} \left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b) \right)$$

Simpson's $3/8$ **rule** $n = 3 \implies A_0 = \frac{3h}{8},\ A_1 = \frac{9h}{8},\ A_2 = \frac{9h}{8},\ A_3 = \frac{3h}{8}$ and

$$\int_a^b f(x)dx = \sum_{i=0}^{3} A_i f(x_i) = \frac{3h}{8}(f(a) + 3f(a+h) + 3f(a+2h) + f(b))$$

Trapezoidal Rule : $n = 1$

Consider interval (a, b). Since $n = 1$, we have $x_0 = a$ and $x_1 = b$.

$$l_0(x) = \frac{x - x_1}{x_0 - x_1}$$

$$l_1(x) = \frac{x - x_0}{x_1 - x_0}$$

$$\begin{aligned} A_0 = \int_a^b l_0(x)dx &= \int_a^b \frac{x - x_1}{x_0 - x_1}dx = \frac{-1}{h}\int_a^b (x - b)dx \\ &= \frac{-1}{h}\left(\frac{(x-b)^2}{2}\right)_a^b = \frac{-1}{h}\left(\frac{0 - (a-b)^2}{2}\right) = \frac{h}{2} \\ A_1 = \int_a^b l_1(x)dx &= \int_a^b \frac{x - x_0}{x_1 - x_0}dx = \frac{1}{h}\int_a^b (x - a)dx \\ &= \frac{1}{h}\left(\frac{(x-a)^2}{2}\right)_a^b = \frac{1}{h}\left(\frac{(b-a)^2 - 0}{2}\right) = \frac{h}{2} \end{aligned}$$

Therefore,

$$\int_a^b f(x)dx = A_0 f(x_0) + A_1 f(x_1) = \frac{h}{2}(f(a) + f(b))$$

Simpon's $1/3$ **Rule :** $n = 2$

Consider interval (a, b) divided into two subintervals of equal length $h = \frac{a+b}{2}$. We have $x_0 = a,\ x_1 = \frac{a+b}{2},\ x_2 = b$.

$$l_0(x) = \frac{x - x_1}{x_0 - x_1}\frac{x - x_2}{x_0 - x_2}$$

$$l_1(x) = \frac{x - x_0}{x_1 - x_0}\frac{x - x_2}{x_1 - x_2}$$

$$l_2(x) = \frac{x - x_0}{x_2 - x_0}\frac{x - x_1}{x_2 - x_1}$$

$$\begin{aligned} A_0 &= \int_a^b l_0(x)dx \\ &= \int_a^b \frac{x - x_1}{x_0 - x_1}\frac{x - x_2}{x_0 - x_2}dx \\ &= \int_a^b \left(\frac{x - \frac{a+b}{2}}{a - \frac{a+b}{2}}\right)\left(\frac{x - b}{a - b}\right)dx \end{aligned}$$

Changing variable of integration $y = x - \frac{a+b}{2}$

$$\begin{aligned} y = x - \frac{a+b}{2} &\implies dy = dx \\ x = a &\implies y = -h \\ x = b &\implies y = h \end{aligned}$$

Continuing with the value of A_0,

$$\begin{aligned} A_0 &= \int_{-h}^{h} \left(\frac{y}{-h}\right)\left(\frac{y - h}{-2h}\right)dy \\ &= \frac{1}{2h^2}\int_{-h}^{h} y^2 - \frac{1}{2h}\int_{-h}^{h} ydy \\ &= \frac{1}{2h^2}\left(\frac{h^3}{3} - \frac{(-h)^3}{3}\right) - \frac{1}{2h}\left(\frac{h^2}{2} - \frac{(-h)^2}{2}\right) \\ &= \frac{h}{3} \end{aligned}$$

$$\begin{aligned} A_1 &= \int_a^b l_1(x)dx \\ &= \int_a^b \frac{x - x_0}{x_1 - x_0}\frac{x - x_2}{x_1 - x_2}dx \\ &= \int_a^b \left(\frac{x - a}{h}\right)\left(\frac{x - b}{-h}\right)dx \end{aligned}$$

Applying change of variable, $y = x - \frac{a+b}{2}$

$$\begin{aligned}
&= \int_{-h}^{h} \left(\frac{y+h}{h}\right)\left(\frac{y-h}{-h}\right) dy \\
&= \frac{-1}{h^2}\int_{-h}^{h} y^2 dy + \int_{-h}^{h} 1 dy \\
&= \frac{-1}{h^2}\left(\frac{h^3}{3} - \frac{(-h)^3}{3}\right) + (h - (-h)) \\
&= \frac{-2h^3}{3h^2} + 2h \\
&= \frac{4h}{3}
\end{aligned}$$

$$\begin{aligned}
A_2 &= \int_a^b l_2(x)dx \\
&= \int_a^b \left(\frac{x-x_0}{x_2-x_0}\right)\left(\frac{x-x_1}{x_2-x_1}\right) dx \\
&= \int_a^b \left(\frac{x-a}{2h}\right)\left(\frac{x-\frac{a+b}{2}}{h}\right) dx
\end{aligned}$$

Applying change of variable, $y = x - \frac{a+b}{2}$

$$\begin{aligned}
&= \int_{-h}^{h} \left(\frac{y+h}{2h}\right)\left(\frac{y}{h}\right) dy \\
&= \frac{1}{2h^2}\int_{-h}^{h} y^2 dy + \frac{1}{2h}\int_{-h}^{h} y dy \\
&= \frac{1}{2h^2}\left(\frac{h^3}{3} - \frac{(-h)^3}{3}\right) + \frac{1}{2h}\left(\frac{h^2}{2} - \frac{(-h)^2}{2}\right) \\
&= \frac{2h^3}{6h^2} \\
&= \frac{h}{3}
\end{aligned}$$

Therefore,

$$\int_a^b f(x)dx = \sum_{i=0}^{2} A_i f(x_i) = \frac{h}{3}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right)$$

14.3.2 Composite Trapezoidal Rule

Suppose an interval (a, b) is divided into n subintervals. In Composite Trapezoidal Rule, Trapezoidal Rule is applied to each subinterval. Thus we have,

$$\begin{aligned}
I &= I_0 + I_1 + \cdots + I_{n-1} \text{ where } I_k \text{ is the integral over } (x_k, x_{k+1}) \\
&= \frac{h(f(x_0) + f(x_1))}{2} + \frac{h(f(x_1) + f(x_2))}{2} + \cdots + \frac{h(f(x_{n-1} + f(x_n))}{2} \\
&= \frac{h}{2}(f(x_0) + 2f(x_1) + \cdots + 2f(x_{n-1}) + f(x_n))
\end{aligned}$$

14.3.3 Recursive Trapezoidal Rule

Suppose an interval (a, b) is divided into 2^{k-1} subintervals. In Recursive Trapezoidal Rule, we apply Trapezoidal Rule on each subinterval. And there is a recursive forumula since the number of intervals doubles as value of k increases by one. Let $H = b - a$

$$\begin{aligned} k = 1 \implies 2^0 = 1 \text{ and} \\ I_1 = &\frac{H}{2}(f(a) + f(b)) \end{aligned}$$

$$\begin{aligned} k = 2 \implies 2^1 = 2 \text{ and} \\ I_2 = &\frac{H}{4}(f(a) + 2f\left(a + \frac{H}{2}\right) + f(b) \\ = &\frac{H}{4}(f(a) + f(b)) + \frac{H}{2}f\left(a + \frac{H}{2}\right) \\ = &\frac{I_1}{2} + \frac{H}{2}f\left(a + \frac{H}{2}\right) \end{aligned}$$

$$\begin{aligned} k = 3 \implies 2^2 = 4 \text{ and} \\ I_3 = &\frac{H}{8}(f(a) + 2f\left(a + \frac{H}{4}\right) + 2f\left(a + \frac{2H}{4}\right) + 2f\left(a + \frac{3H}{4}\right) + f(b)) \\ = &\frac{H}{8}(f(a) + 2f\left(a + \frac{2H}{4}\right) + 2f\left(a + \frac{4H}{4}\right) + 2f\left(a + \frac{6H}{4}\right) + f(b)) \\ &+ \frac{H}{4}(f\left(a + \frac{H}{4}\right) + f\left(a + \frac{3H}{4}\right)) \\ = &\frac{I_2}{2} + \frac{H}{4}(f\left(a + \frac{H}{4}\right) + f\left(a + \frac{3H}{4}\right)) \end{aligned}$$

Consider interval $(0, 64)$. We have $b - a = H = 64$.
$I_1 = 32(f(0) + f(64))$
$I_2 = 16(f(0) + 2f(32) + f(64))$
$I_3 = 8(f(0) + 2f(16) + 2f(32) + 2f(48) + f(64))$
$I_4 = 4(f(0) + 2f(8) + 2f(16) + \cdots + 2f(56) + f(64)$
$I_5 = 2(f(0) + 2f(4) + 2f(8) + \cdots + 2f(60) + f(64)$
$I_6 = f(0) + 2f(2) + 2f(4) + \cdots + 2f(62) + f(64)$
$\vdots$
The values corresponding to the intervals in I_{k-1} appear in I_k as alternate terms. Other terms, corresponds to the odd multiples of $\frac{H}{2^k}$. We separate them into two sums and represent the first sum as $\frac{I_{k-1}}{2}$.

$$\begin{aligned} \text{Clearly, } I_k = &\frac{H}{2^k}\sum_{i=0}^{2^{k-2}} f\left(a + \frac{2iH}{2^k}\right) + \frac{2H}{2^k}\sum_{i=1}^{2^{k-2}} f\left(a + \frac{(2i-1)H}{2^{k-1}}\right) \\ = &\frac{I_{k-1}}{2} + \frac{H}{2^{k-1}}\sum_{i=1}^{2^{k-2}} f\left(a + \frac{(2i-1)H}{2^{k-1}}\right) \end{aligned}$$

14.3.4 Python : Recursive Trapezoidal Rule

Program 14.3.

```
def recursiveTrapezoidalRule(f,a,b,Iold,k):
  if k == 1 :
    Inew = (f(a)+f(b))*(b-a)/2.0
  else :
    n = 2**(k-2)
    h = (b-a) * 1.0 / n
    x = a + h/2.0
    sum = 0.0
    for i in range(n):
      sum = sum + f(x)
      x = x + h
    Inew = (Iold + h * sum ) / 2.0
  return Inew
```

Line 1 `def recursiveTrapezoidalRule(f,a,b,Iold,k):`
This function has five input arguments. (a) f is a real function (b) a,b are start and end of interval in which f is going to be integrated (c) $Iold$ is the value of the integral for 2^{k-1} subintevals using recursive trapezoidal method (d) k is a variable such that 2^k is the number of subintervals considered for Integration.

Line 2 `if` $k == 1$ `:`
If $k = 1$, we proceed to Line 3, otherwise we go to Line 4.

Line 3 $Inew = (f(a) + f(b)) * (b - a)/2.0$
For $k = 1$, we use trapezoidal rule $I_1 = \frac{b-a}{2}(f(a) + f(b))$. When writing this in python, we are use 2.0 so that the python won't ignore the decimal part of this fraction. In python, $5/2 = 2$. And $5/2.0 = 2.5$

Line 4 `else :`
If Line 2 is false, (ie $k \neq 1$) python executes Line 5-8. These line implements the recursive formula for I_k.

Line 5 $n = 2 * *(k - 2)$
Equivalent to $n \leftarrow 2^{k-2}$.

Line 6 $h = (b - a) * 1.0/n$
This the length of a subinterval when we divide (a, b) into 2^k subintervals/panels. Equivalent to $h \leftarrow \frac{b-a}{2^{k-2}}$

Line 7 $x = a + h/2.0$
This the parameter of f in the first term in the sum $\sum_{i=1}^{2^{k-2}} f\left(a + \frac{(2i-1)H}{2^{k-1}}\right)$ in the recursive formula for I_k. Equivalent to $x \leftarrow a + \frac{h}{2}$.

Line 8 $sum = 0.0$
We are going to use this variable to find that sum. To start with, we will make it 0 and will add each term to it one-by-one. Equivalent to $sum \leftarrow 0$.

Line 9 `for` i `in range(`n`):`
This the variable i in the recursive forumula for I_k. For each value of

$i = 1, 2, \cdots, 2^{k-2}$, Lines 10 and 11 are executed. That is, for each value of i, the corresponding term in the sum is computed and added to the variable *sum*.

Line 10 $sum = sum + f(x)$
Value of f at x is computed and added to the partial sum. Equivalent to $sum \leftarrow sum + f(x)$. However, the value of x is changed for each i in Line 11. Thus, for next value of i, x and *sum* have the new values to use.

Line 11 $x = x + h$
Equivalent to $x \leftarrow x + h$. For $i = 1$, $x = a + \frac{h}{2}$ before Line 11. At Line 11, $x \leftarrow a + \frac{3h}{2}$. And this is the value of x for $i = 2$ before Line 11 next time. Thus, x iterates through $a + \frac{h}{2}, a + \frac{3h}{2}, \cdots, a + \frac{(2n-1)h}{2}$. This x is updated and used for next execution of Line 10 and 11.

Line 12 $Inew = (Iold + h * sum)/2.0$
We reach here only after executing Line 10-11 for all values of i. That is, *sum* in the recursive forumula is already computed. This line, implements the recursive formula and stores that value into the variable $Inew$. Equivalent to $Inew \leftarrow \frac{Iold + h \times sum}{2}$.

Line 13 `return Inew`
It returns the value of I_k, the intergral of f over (a, b) using recursive trapezoidal rule for 2^k subintervals.

Part III

ME010303 Multivariate Calculus & Integral Transforms

Chapter 15

Fourier Series and Fourier Integrals

15.1 The Weierstrass Approximation Theorem

Every continuous, real valued function on a compact interval has a polynomial approximation.[Apostol, 1973, Theorem 11.17]

Theorem 15.1 (Weierstrass)**.** *Let f be a real-valued, continuous function on a compact interval $[a, b]$. Then for every $\epsilon > 0$, there is a polynomial p such that $|f(x) - p(x)| < \epsilon$ for every $x \in [a, b]$.*

Synopsis. *Given a real-valued continuous function on compact interval $[a, b]$, we can construct a real-valued, continous function g on $\mathbb{R}$ which is periodic with period 2π. We have, if $f \in L(I)$ and f is bounded almost everywhere in I, then $f \in L^2(I)$.[Apostol, 1973, Theorem 10.52].By Fejer's theorem ([Apostol, 1973, Theorem 11.15]), the fourier series generated by g ([Apostol, 1973, definition 11.3]) converges to the Cesaro sum ([Apostol, 1973, Definition 8.47]), which is g itself in this case. Thus for any $\epsilon > 0$, there is a finite sum of trignometric functions. The power series expansions of trignometric functions ([Apostol, 1973, definition 9.27]) being uniformly convergent, there exists a polynomial p_m which approximates g. And we can construct p (polynomial approximation of g) using p_m.*

Proof. Define $g : \mathbb{R} \to \mathbb{R}$,

$$g(t) = \begin{cases} f(a + (b-a)t/\pi), \ t \in [0, \pi) \\ f(a + (2\pi - t)(b-a)/\pi), \ t \in [\pi, 2\pi] \\ g(t - 2n\pi), \ t > 2\pi, \ n \in \mathbb{N} \\ g(t + 2n\pi), \ t < 0, \ n \in \mathbb{N} \end{cases}$$

Thus g is a continuous, real-valued, periodic function with period 2π such that

$$f(x) = g\left(\frac{\pi(x-a)}{b-a}\right), \ x \in [a, b] \tag{15.1}$$

The fourier series generated by g is given by,

$$g(t) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt)$$

$$\text{where } a_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos kt\ dt,\ b_k = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin kt\ dt$$

Let $\{s_n(t)\}$ be the sequence of partial sums of the fourier series generated by g. And $\{\sigma_n(t)\}$ be the sequence of averages of $s_n(t)$ given by,

$$\sigma_n(t) = \frac{1}{n}\sum_{k=1}^{n} s_k(t),\ \text{where } s_k(t) = \frac{a_0}{2} + \sum_{j=1}^{k} (a_j \cos jt + b_j \sin jt)$$

Function $f \in L(I)$ being real-valued continuous function on a compact interval, it is bounded and hence is Lebesgue square integrable.ie, $f \in L^2(I)$. Thus, $g \in L^2(I)$.

Since g is continous on $\mathbb{R}$, the function $s : \mathbb{R} \to \mathbb{R}$ defined by,

$$s(t) = \lim_{h\to 0^+} \frac{g(t+h) - g(t-h)}{2}$$

is well-defined on $\mathbb{R}$ and $s(t) = g(t),\ \forall t \in \mathbb{R}$.

Then by Fejer's Theorem, the sequence $\{\sigma_n(t)\}$ converges uniformly to $g(t)$ for every $t \in \mathbb{R}$. Thus, given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $\forall t \in \mathbb{R}$, $|g(t) - \sigma_N(t)| < \frac{\epsilon}{2}$.

We have,

$$\sigma_N(t) = \sum_{k=0}^{N} (A_k \cos kt + B_k \sin kt)\,,\ \text{where } A_k, B_k \in \mathbb{R} \tag{15.2}$$

By the power series expansion of the trignometric functions about origin,

$$\cos kt = \sum_{j=1}^{\infty} \left(\frac{\cos^{(j)} 0}{j!}(kt)^j\right) = \sum_{j=1}^{\infty} A'_j t^j \text{ where } A'_j \in \mathbb{R} \tag{15.3}$$

$$\sin kt = \sum_{j=1}^{\infty} \left(\frac{\sin^{(j)} 0}{j!}(kt)^j\right) = \sum_{j=1}^{\infty} B'_j t^j \text{ where } B'_j \in \mathbb{R} \tag{15.4}$$

Since the above power series expansions of trignometric functions are uniformly convergent, their finite linear combination $\{\sigma_N(t)\}$ is also uniformly convergent. ie, Given $\epsilon > 0$ there exists $m \in \mathbb{N}$ such that for every $t \in \mathbb{R}$

$$\left|\sum_{k=0}^{m} C_k t^k - \sigma_N(t)\right| < \frac{\epsilon}{2} \text{ where } C_k \in \mathbb{R}$$

Therefore, $|p_m(t) - g(t)| \le |p_m(t) - \sigma_N(t)| + |\sigma_N(t) - g(t)| < \epsilon$ where $p_m(t) = \sum_{k=0}^{m} C_k t^k$. Define $p : [a, b] \to \mathbb{R}$ by,

$$p(x) = p_m\left(\frac{\pi(x-a)}{b-a}\right) \tag{15.5}$$

By equations 15.1 and 15.5, $|p(x) - f(x)| < \epsilon$ for every $x \in [a, b]$. □

15.2 Other Forms of Fourier Series

Let $f \in L([0, 2\pi])$, then the fourier series generated by f is given by,

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

$$\text{where } a_n = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos nt \; dt, \qquad b_n = \frac{1}{\pi} \int_0^{2\pi} f(t) \sin nt \; dt$$

By Euler's forumula $e^{inx} = \cos nx + i \sin nx$. We have, $\cos nx = \frac{(e^{inx} + e^{-inx})}{2}$ and $\sin nx = \frac{(e^{inx} - e^{-inx})}{2i}$

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(\alpha_n e^{inx} + \beta_n e^{-inx}\right)$$

$$\text{where } \alpha_n = \frac{(a_n - ib_n)}{2} \qquad \beta_n = \frac{(a_n + ib_n)}{2}$$

Therefore, by assigning $\alpha_0 = a_0/2$, $\alpha_{-n} = \beta_n$, we get the following exponential form of fourier series generated by f,

$$f(x) \sim \sum_{n=-\infty}^{\infty} \alpha_n e^{inx} \text{ where } \alpha_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) \; e^{-int} \; dt$$

Note : If f is periodic with period 2π, then the interval of integration $[0, 2\pi]$ can be replaced with any interval of length 2π. eg. $[-\pi, \pi]$

15.2.1 Periodic with period p

Let $f \in L([0, p])$ and f is periodic with period p. Then

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{2\pi nx}{p} + b_n \sin \frac{2\pi nx}{p}\right)$$

$$\text{where } a_n = \frac{2}{p} \int_0^p f(t) \cos \frac{2\pi nt}{p} \; dt \qquad b_n = \frac{2}{p} \int_0^p f(t) \sin \frac{2\pi nt}{p} \; dt$$

Therefore, we have the exponential form of the above fourier series given by,

$$f(x) \sim \sum_{n=-\infty}^{\infty} \alpha_n e^{\frac{2\pi inx}{p}}, \text{ where } \alpha_n = \frac{1}{p} \int_0^p f(t) \; e^{\frac{-2\pi int}{p}} \; dt$$

15.3 Fourier Integral Theorem

Theorem 15.2 (Fourier Integral Theorem)**.** *Let $f \in L(-\infty, \infty)$. Suppose $x \in \mathbb{R}$ and an interval $[x - \delta, x + \delta]$ about x such that either*

1. *f is of bounded variation on an interval $[x - \delta, x + \delta]$ about x or*

2. both limits $f(x+)$ and $f(x-)$ exists and both Lebesgue intergrals

$$\int_0^\delta \frac{f(x+t)-f(x+)}{t}dt \text{ and } \int_0^\delta \frac{f(x-t)-f(x-)}{t}dt$$

exists.

Then,

$$\frac{f(x+)+f(x-)}{2} = \frac{1}{\pi}\int_0^\infty \int_{-\infty}^\infty f(u)\cos v(u-x)\ du\ dv,$$

the integral $\int_0^\infty$ being an improper Riemann integral.

Synopsis.

$$f(x+t)\frac{\sin \alpha t}{\pi t}dt \to f(u)\frac{\sin \alpha(u-x)}{\pi(u-x)} \to \frac{f(u)}{\pi}\int_0^\alpha \cos v(u-x)dv$$

By Riemann-Lebesgue lemma[Apostol, 1973, Theorem 11.6],

$$f \in L(I) \implies \lim_{\alpha\to+\infty}\int_I f(x)\sin \alpha t\ dt = 0$$

By Jordan's Theorem[Apostol, 1973, Theorem 10.8], if g is of bounded variation on $[0,\delta]$, then

$$\lim_{\alpha\to+\infty}\frac{2}{\pi}\int_0^\delta g(t)\frac{\sin \alpha t}{t}dt = g(0+)$$

By Dini's Theorem[Apostol, 1973, Theorem 10.9], if the limit $g(x+)$ exists and Lebesgue integral $\int_0^\delta \frac{g(t)+g(0+)}{t}dt$ exists for some $\delta > 0$, then

$$\lim_{\alpha to+\infty}\frac{2}{\pi}\int_0^\delta g(t)\frac{\sin \alpha t}{t}dt = g(0+)$$

The order of Lebesgue integrals can be interchanged[Apostol, 1973, Theorem 10.40],
Suppose $f \in L(X)$ and $g \in L(Y)$. Then

$$\int_X f(x)\left(\int_Y g(y)k(x,y)dy\right)dx = \int_Y g(y)\left(\int_X f(x)k(x,y)dx\right)dy$$

Proof. Consider $\int_{-\infty}^\infty f(x+t)\frac{\sin \alpha t}{\pi t}dt$. We prove that this integral is equal to the either sides.

$$\int_{-\infty}^\infty f(x+t)\frac{\sin \alpha t}{\pi t}dt = \int_{-\infty}^{-\delta} + \int_{-\delta}^0 + \int_0^{-\delta} + \int_\delta^\infty f(x+t)\frac{\sin \alpha t}{\pi t}dt$$

We have, function $\frac{f(x+t)}{\pi t}$ is bounded on $(-\infty,-\delta)\cup(\delta,\infty)$, hence $\frac{f(x+t)}{\pi t}$ is Lebesgue integrable on $(-\infty,-\delta)\cup(\delta,\infty)$.

By Riemann Lebesgue lemma,

$$\frac{f(x+t)}{\pi t} \in L(-\infty,-\delta) \implies \int_{-\infty}^{-\delta} f(x+t)\frac{\sin \alpha t}{\pi t}dt = 0,$$

$$\frac{f(x+t)}{\pi t} \in L(\delta,\infty) \implies \int_\delta^\infty f(x+t)\frac{\sin \alpha t}{\pi t}dt = 0$$

Case 1 Suppose f is of bounded variation on $[x-\delta, x+\delta]$, put $g(t) = f(x+t)$ then g is of bounded variation on $[-\delta, \delta]$. Thus g is of bounded variation on $[0, \delta]$. Then by Jordan's Theorem

$$\lim_{\alpha\to+\infty} \frac{2}{\pi}\int_0^\delta f(x+t)\frac{\sin\alpha t}{t}dt = \lim_{\alpha\to+\infty} \frac{2}{\pi}\int_0^\delta g(t)\frac{\sin\alpha t}{t}dt = g(0+) = f(x+)$$

Case 2 Suppose both the limits $f(x+)$ and $f(x-)$ exists and both Lebesgue integrals

$$\int_0^\delta \frac{f(x+t)-f(x+)}{t}dt \text{ and } \int_0^\delta \frac{f(x-t)-f(x-)}{t}dt$$

exists.

Thus, we have $f(x+)$ exists and the Lebesgue integral $\int_0^\delta \frac{f(x+t)-f(x+)}{t}dt$ exists. Put $g(t) = f(x+t)$, then $g(0+) = f(x+)$ exists and the Lebesgue integral $\int_0^\delta \frac{g(t)-g(0+)}{t}dt$ exists, then by Dini's Theorem,

$$\lim_{\alpha\to+\infty} \frac{2}{\pi}\int_0^\delta f(x+t)\frac{\sin\alpha t}{t}dt = \lim_{\alpha\to+\infty} \frac{2}{\pi}\int_0^\delta g(t)\frac{\sin\alpha t}{t}dt = g(0+) = f(x+)$$

Similarly, $f(x-)$ exists and the Lebesgue integral $\int_0^\delta \frac{f(x-t)-f(x-)}{t}dt$ exists. Put $g(t) = f(x-t)$, then $g(0+) = f(x-)$ exists and the Lebesgue integral $\int_0^\delta \frac{g(t)-g(0+)}{t}dt$ exists, then by Dini's Theorem,

$$\begin{aligned}\lim_{\alpha\to+\infty} \frac{2}{\pi}\int_{-\delta}^0 f(x+t)\frac{\sin\alpha t}{t}dt &= \lim_{\alpha\to+\infty} \frac{2}{\pi}\int_0^\delta f(x-\tau)\frac{\sin\alpha\tau}{\tau}d\tau \\ &= \lim_{\alpha\to+\infty} \frac{2}{\pi}\int_0^\delta g(\tau)\frac{\sin\alpha\tau}{\tau}d\tau = g(0+) = f(x-)\end{aligned}$$

Then by either cases,

$$\begin{aligned}\lim_{\alpha\to+\infty}\int_{-\infty}^\infty f(x+t)\frac{\sin\alpha t}{\pi t}dt &= \lim_{\alpha\to+\infty}\int_{-\delta}^0 + \int_0^\delta f(x+t)\frac{\sin\alpha t}{\pi t}dt \\ &= \frac{f(x+)+f(x-)}{2}\end{aligned}$$

We have, $\int_0^\alpha \cos v(u-x)dv = \frac{\sin v(u-x)}{u-x}$.

$$\begin{aligned}\lim_{\alpha\to+\infty}\int_{-\infty}^\infty f(x)\frac{\sin\alpha t}{\pi t}dt &= \lim_{\alpha\to+\infty}\int_{-\infty}^\infty f(u)\frac{\sin\alpha(u-x)}{u-x}du, \text{ (put } u = x+t) \\ &= \lim_{\alpha\to+\infty}\int_{-\infty}^\infty f(u)\left(\int_0^\alpha \cos v(u-x)dv\right)du \\ &= \lim_{\alpha\to+\infty}\int_0^\alpha \left(\int_{-\infty}^\infty f(u)\cos v(u-x)du\right)dv, \\ &\quad \text{since, the order of Lebesgue integrals can be reversed.} \\ &= \int_0^\infty \left(\int_{-\infty}^\infty f(u)\cos v(u-x)du\right)dv\end{aligned}$$

where, $\int_0^\infty$ is not a Lebesgue integral, but an improper Riemann integral

Therefore,

$$\int_0^\infty \left(\int_{-\infty}^\infty f(u) \cos v(u-x) du \right) dv = \lim_{\alpha \to +\infty} \int_{-\infty}^\infty f(x) \frac{\sin \alpha t}{\pi t} dt$$
$$= \frac{f(x+) + f(x-)}{2}$$

□

Remark. *If a function f on $(-\infty, \infty)$ is non-periodic, then it may not have a fourier series represenation. In such cases, we have fourier intergral representaion.*

15.4 Exponential form of Fourier Integral Theorem

Let $f \in L(-\infty, \infty)$. Suppose $x \in \mathbb{R}$ and an interval $[x-\delta, x+\delta]$ about x such that either

1. f is of bounded variation on an interval $[x-\delta, x+\delta]$ about x or
2. both limits $f(x+)$ and $f(x-)$ exists and both Lebesgue intergrals

$$\int_0^\delta \frac{f(x+t) - f(x+)}{t} dt \text{ and } \int_0^\delta \frac{f(x-t) - f(x-)}{t} dt$$

exists.

Then,

$$\frac{f(x+) + f(x-)}{2} = \lim_{\alpha \to \infty} \frac{1}{2\pi} \int_{-\alpha}^\alpha \left(\int_{-\infty}^\infty f(u) e^{iv(u-x)} \, du \right) dv$$

Proof. Let $F(v) = \int_{-\infty}^\infty f(u) \cos v(u-x) du$. Then $F(v) = F(-v)$ and

$$\lim_{\alpha \to \infty} \frac{1}{2\pi} \int_{-\alpha}^\alpha F(v) dv = \lim_{\alpha \to \infty} \frac{1}{\pi} \int_0^\alpha \int_{-\infty}^\infty f(u) \cos v(u-x) du dv$$
$$= \frac{f(x+) + f(x-)}{2}$$

Let $G(v) = \int_{-\infty}^\infty f(u) \sin v(u-x) du$. Then $G(v) = -G(-v)$ and

$$\lim_{\alpha \to \infty} \frac{1}{2\pi} \int_{-\alpha}^\alpha G(v) dv = 0$$

Thus

$$\lim_{\alpha \to \infty} \frac{1}{2\pi} \int_{-\alpha}^\alpha F(v) + iG(v) dv = \frac{f(x+) + f(x-)}{2}$$

□

15.5 Integral Transforms

Definitions 15.3. *Integral transform $g(y)$ of $f(x)$ is a Lebesgue integral or Improper Riemann integral of the form*

$$g(y) = \int_{-\infty}^{\infty} K(x,y)f(x)\ dx$$

, where K is the kernal of the transform. We write $g = \mathscr{K}(f)$.

Remark. *Integral transforms(operators) are linear operators.*
ie, $\mathscr{K}(af_1 + bf_2) = a\mathscr{K}f_1 + b\mathscr{K}f_2$

Remark. *A few commonly used integral transforms,*

1. *Exponential Fourier Transform $\mathscr{F}$,*

$$\mathscr{F}f = \int_{-\infty}^{\infty} e^{-ixy}f(x)\ dx$$

2. *Fourier Cosine Transform $\mathscr{C}$,*

$$\mathscr{C}f = \int_{0}^{\infty} \cos xyf(x)\ dx$$

3. *Fourier Sine Transform $\mathscr{S}$,*

$$\mathscr{S}f = \int_{0}^{\infty} \sin xyf(x)\ dx$$

4. *Laplace Transform $\mathscr{L}$,*

$$\mathscr{L}f = \int_{0}^{\infty} e^{-xy}f(x)\ dx$$

5. *Mellin Transform $\mathscr{M}$,*

$$\mathscr{M}f = \int_{0}^{\infty} x^{y-1}f(x)\ dx$$

Remark. *Suppose $f(x) = 0,\ \forall x < 0$.*

$$\int_{-\infty}^{\infty} e^{-ixy}f(x)\ dx = \int_{0}^{\infty} e^{-ixy}f(x)\ dx = \int_{0}^{\infty} \cos xy\ f(x)\ dx + i\int_{0}^{\infty} \sin xy\ f(x)\ dx$$

$$\mathscr{F}f = \mathscr{C}f + i\mathscr{S}f$$

Therefore Fourier Cosine $\mathscr{C}$ and Sine $\mathscr{S}$ transforms are special cases of fourier integral transform, $\mathscr{F}$ provided f vanishes on negative real axis.

Remark. *Let $y = u + iv$, $f(x) = 0$, $\forall x < 0$.*

$$\int_0^\infty e^{-xy} f(x) = \int_0^\infty e^{-xu} e^{-ixv} f(x)\ dx = \int_0^\infty e^{-ixv} \phi_u(x) dx$$

where $\phi_u(x) = e^{-xu} f(x)$.

$$\mathscr{L} f = \mathscr{F} \phi_u$$

Therefore Laplace transform, $\mathscr{L}$ is a special case of Fourier integral transform, $\mathscr{F}$.

Remark. *Let $g(y) = \mathscr{F} f(x)$.*

$$g(y) = \int_{-\infty}^\infty e^{-ixy} f(x)\ dx$$

Suppose f is continuous at x, then by fourier integral theorem,

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^\infty \left(\int_{-\infty}^\infty f(u) e^{iv(u-x)} du \right) dv \\ &= \int_{-\infty}^\infty e^{-ivx} \left(\frac{1}{2\pi} \int_{-\infty}^\infty e^{ivu} f(u)\ du \right) dv \\ &= \int_{-\infty}^\infty g(v) e^{-ivx} dv = \mathscr{F} g \text{ where } g(v) = \frac{1}{2\pi} \int_{-\infty}^\infty f(u) e^{ivu} du \end{aligned}$$

The above function $g(v)$ gives the ***inverse fourier transformation*** *of f.*

Let g be fourier transform of f, then f is uniquely determined by its fourier transform g by,

$$f(x) = \mathscr{F}^{-1} g(y) = \frac{1}{2\pi} \lim_{\alpha \to \infty} \int_{-\alpha}^{\alpha} g(y) e^{ixy} dy$$

6. Inverse Fourier Transform $\mathscr{F}^{-1}$,

$$\mathscr{F}^{-1} f = \int_{-\infty}^\infty \frac{e^{ixy}}{2\pi} f(x)\ dx$$

15.6 Convolutions

Definitions 15.4. *Let $f, g \in L(-\infty, \infty)$. Let S be the set of all points x for which the Lebesgue integral*

$$h(x) = \int_{-\infty}^\infty f(t) g(x-t) dt$$

exists. Then the function $h : S \to \mathbb{R}$ is a convolution of f and g. And $h = f * g$.

Remark. *Convolution operator is commutative.*
ie, $h = f * g = g * f$

Remark. *Suppose f, g vanishes on negative real axis, then*

$$h(x) = \int_{-\infty}^\infty f(t)\ g(x-t)\ dt = \int_{-\infty}^0 + \int_0^x + \int_x^\infty f(t)\ g(x-t)\ dt = \int_0^x f(t)\ g(x-t)\ dt$$

Remark. *Singularity is a point at which the convolution integral fails to exists.*

Theorem 15.5. *Let $f, g \in L(\mathbb{R})$ and either f or g is bounded in $\mathbb{R}$. Then the convoluton integral*

$$h(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt$$

exists for every $x \in \mathbb{R}$ and the function h so defined is bouned in $\mathbb{R}$. In addition, if the bounded function is continuous on $\mathbb{R}$, then h is continuous and $h \in L(\mathbb{R})$.

Synopsis.

Proof. □

Remark. *If f, g are both unbounded, the convolution integral may not exist.*

$$\text{eg: } f(t) = \frac{1}{\sqrt{t}},\ g(t) = \frac{1}{\sqrt{1-t}}$$

Theorem 15.6. *Let $f, g \in L^2(\mathbb{R})$. Then the convolution integral $f * g$ exists for each $x \in \mathbb{R}$ and the function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x) = f * g(x)$ is bounded in $\mathbb{R}$.*

Synopsis.

Proof. □

15.7 The Convolution Theorem for Fourier Tranforms

Theorem 15.7. *Let $f, g \in L(\mathbb{R})$ and either f or g is continuous and bounded on $\mathbb{R}$. Let $h = f * g$. Then for every real u,*

$$\int_{-\infty}^{\infty} h(x)e^{-ixu}dx = \left(\int_{-\infty}^{\infty} f(t)e^{-itu}dt\right)\left(\int_{-\infty}^{\infty} g(y)e^{-iyu}dy\right)$$

The integral on the left exists both as a Lebesgue integral and an improper Riemann integral.

Synopsis.

Proof. □

Remark (Application of Convolution Theorem)**.**

$$B(p,q) = \frac{\Gamma p \Gamma q}{\Gamma p + q},\ \textit{where } B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx,\ \Gamma p = \int_0^{\infty} t^{p-1}e^{-t}dt$$

Chapter 16

Multivariate Differential Calculus

In this chapter, we deal with real functions of several variables. Instead of $\mathbf{c}$, we write $\overline{c} \in \mathbb{R}^n$, then $\overline{c} = (c_1, c_2, \ldots, c_n)$ where $c_j \in \mathbb{R}$ for every $j = 1, 2, \ldots, n$. Again, suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ and $f(\overline{x}) = \overline{y}$, then $\overline{y} = (y_1, y_2, \ldots, y_m)$ where each y_k is real. The unit co-ordinate vector, $\overline{u_k}$ is given by $u_{kj} = \delta_{j,k}$

16.1 Directional Derivative

Motivation : The existence of all partial derivatives of a multivariate real function f at a point $\overline{c}$ doesn't imply the continuity of f at $\overline{c}$. Thus, we need a suitable generalisation for the partial derivative which could characterise continuity. And directional derivative is such an attempt.

Definitions 16.1 (Directional Derivative)**.** *Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}^m$. Let $\overline{c}$ be an interior points of S and $\overline{u} \in \mathbb{R}^n$, then there exists an open ball $B(\overline{c}, r)$ in S. Also for some $\delta > 0$ the line segment $\alpha : [0, \delta] \to S$ given by $\alpha(t) = \overline{c} + t\overline{u}$ lie in $B(\overline{c}, r)$.*

Then the Directional derivative of f at an interior point $\overline{c}$ in the direction $\overline{u}$ is given by

$$f'(\overline{c}, \overline{u}) = \lim_{h \to 0} \frac{f(\overline{c} + h\overline{u}) - f(\overline{c})}{h}$$

Remark. *The direction derivative of f at an interior point $\overline{c}$ in the direction $\overline{u}$ exists only if the above limit exists.*

Remark. *Example, [Apostol, 1973, Exercise 12.2a]*
Suppose $\overline{x}, \overline{a}, \overline{c}, \overline{u} \in \mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ such that $f(\overline{x}) = \overline{a} \cdot \overline{x}$. Then

$$f'(\overline{c}, \overline{u}) = \lim_{h \to 0} \frac{\overline{a} \cdot (\overline{c} + h\overline{u}) - \overline{a} \cdot \overline{c}}{h} = \overline{a} \cdot \overline{u}$$

Remark (Properties)**.** *Let $f : S \to \mathbb{R}^m$, where $S \subset \mathbb{R}^n$*

1. $f'(\overline{c}, \overline{0}) = \overline{0}$
   Note : The zero vectors belongs to $\mathbb{R}^n, \mathbb{R}^m$ respectively.

2. $f'(\overline{c}, \overline{u_k}) = \frac{\partial f}{\partial u_k}(\overline{c}) = D_k f(\overline{c})$, *the* k^{th} *partial derivative of* f.

3. *Let* $f = (f_1, f_2, \cdots, f_m)$, *such that* $f(\overline{c}) = (f_1(\overline{c}), f_2(\overline{c}), \ldots, f_m(\overline{c}))$. *Then,*

   $$\exists f'(\overline{c}, \overline{u}) \iff \forall k, \exists f'_k(\overline{c}, \overline{u}) \text{ and } f'(\overline{c}, \overline{u}) = (f'_1(\overline{c}, \overline{u}), f'_2(\overline{c}, \overline{u}), \ldots, f'_m(\overline{c}, \overline{u}))$$

   ie, Directional derivative of f *exists iff directional derivative of each component function* f_k *exists. And the components of the directional derivatives of* f *are the directional derivaties of the components of* f.
   Thus $D_k f(\overline{c}) = (D_k f_1(\overline{c}), D_k f_2(\overline{c}), \ldots, D_k f_m(\overline{c}))$ *holds.*

4. *Let* $F(t) = f(\overline{c} + t\overline{u})$, *then* $F'(0) = f'(\overline{c}, \overline{u})$ *and* $F'(t) = f'(\overline{c} + t\overline{u}, \overline{u})$

5. *Let* $f(\overline{c}) = \overline{c} \cdot \overline{c} = \|\overline{c}\|^2$, *and* $F(t) = f(\overline{c} + t\overline{u})$, *then* $F'(t) = 2\overline{c} \cdot \overline{u} + 2t\|\overline{u}\|^2$ *and* $F'(0) = f'(\overline{c}, \overline{u}) = 2\overline{c} \cdot \overline{u}$

6. *Let* f *be linear, then* $f'(\overline{c}, \overline{u}) = f(\overline{u})$

7. *Existence of all partial derivatives doesn't imply existence of all directional derivatives.*

   $$f(x, y) = \begin{cases} x + y & \text{if } x = 0 \text{ or } y = 0 \\ 1 & \text{otherwise} \end{cases}$$

   For above f, *directional derivatives exists only along the co-ordinates (ie, partial derivatives).*

8. *Existence of all directional derivatives doesn't imply continuity.*

   $$f(x, y) = \begin{cases} xy^2(x^2 + y^4) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

   Above f *is discontinuous at* $(0, 0)$, *however all directional derivatives exists and has finite value.*

16.2 Total Derivative

We may define a total derivative $T_c(h) = hf'(c)$ in the case of real-functions of single variable as follows :-

$$\text{Let } E_c(h) = \begin{cases} \frac{f(c+h)-f(c)}{h} - f'(c), & h \neq 0 \\ 0, & h = 0 \end{cases}$$

Then, $f(c + h) = f(c) + hf'(c) + hE_c(h)$ and as $h \to 0$, $E_c(h) \to 0$. Also $T_c(h) = f'(c)h$ is a linear function of h. ie, $T_c(ah_1 + bh_2) = aT_c(h_1) + bT_c(h_2)$. Now, we will define a total derivative of multivariate function that has these two properties.

Definitions 16.2 (Total Derivative)**.** *The function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\overline{c}$ *if there exists a* ***linear*** *function* $T_{\overline{c}} : \mathbb{R}^n \to \mathbb{R}^m$ *such that* $f(\overline{c} + \overline{v}) = f(\overline{c}) + T_{\overline{c}}(\overline{v}) + \|\overline{v}\|E_{\overline{c}}(\overline{v})$ *where* $E_{\overline{c}}(\overline{v}) \to \overline{0}$ *as* $\overline{v} \to \overline{0}$.

The linear function $T_{\overline{c}}$ is the total derivative of f at $\overline{c}$, $T_{\overline{c}}(\overline{0}) = \overline{0}$ and the condition above gives the First Order Taylor's Formula for $f(\overline{c}+\overline{v}) - f(\overline{c})$.

Remark (Properties)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $f'(\overline{c})(\overline{v}) = T_{\overline{c}}(\overline{v})$ be the total derivative of f at $\overline{c}$ evaluated at $\overline{v}$. Then,*

1. $f'(\overline{c})(\overline{v}) = f'(\overline{c}, \overline{u})$
2. *If f is differentiable at $\overline{c}$, then f is continuous at $\overline{c}$.*
3. $f'(\overline{c})(\overline{v}) = v_1 D_1 f(\overline{c}) + v_2 D_2 f(\overline{c}) + \cdots + v_n D_n f(\overline{c})$

Note. *The above f' is a function from $\mathbb{R}^n$ to the set of all linear functions $\mathscr{L} = \{h : \mathbb{R}^n \to \mathbb{R}^m\}$. $f'(\overline{c})$ is a linear function (in fact, total derivative $T_{\overline{c}}$) which maps $\overline{v}$ into the directional derivatives of f at $\overline{c}$ in the direction $\overline{v}$. This notation generalises f' for univariate f as well.(put $n = m = 1$)*

In this subject, we use the following notations,

$D_k f(\overline{c})$ *partial derivative*

$f'(\overline{c}, \overline{v})$ *directional derivative*

$f'(\overline{c})(\overline{v})$ *total derivative*

$\nabla f(\overline{c})$ *gradient vector*

Theorem 16.3. *If f is differentiable at $\overline{c}$ with total derivative $T_{\overline{c}}$, then for every $\overline{u} \in \mathbb{R}^n$, $T_{\overline{c}}(\overline{u}) = f'(\overline{c}, \overline{u})$. (ie, $f'(\overline{c})(\overline{v}) = f'(\overline{c}, \overline{v})$)*

Proof. For $\overline{v} = \overline{0}$, we have $T_{\overline{c}}(\overline{0}) = 0 = f'(\overline{c}, \overline{0})$.

Suppose $\overline{v} \neq \overline{0}$, then put $\overline{v} = h\overline{u}$. Since f is differentiable at $\overline{c}$, f has total derivative at $\overline{c}$. That is, there exists a linear function $T_{\overline{c}}$ such that $f(\overline{c}+h\overline{u}) = f(\overline{c}) + T_{\overline{c}}(h\overline{u}) + \|h\overline{u}\| E_{\overline{c}}(h\overline{u})$ where $E_{\overline{c}}(h\overline{u}) \to \overline{0}$ as $h\overline{u} \to \overline{0}$.

$$
\begin{aligned}
&\implies f(\overline{c}+h\overline{u}) = f(\overline{c}) + hT_{\overline{c}}(\overline{u}) + |h|\|\overline{u}\| E_{\overline{c}}(h\overline{u}),\ E_{\overline{c}}(h\overline{u}) \to \overline{0} \text{ as } h\overline{u} \to \overline{0} \\
&\implies \frac{f(\overline{c}+h\overline{u}) - f(\overline{c})}{h} = T_{\overline{c}}(\overline{u}) + \frac{|h|\|\overline{u}\| E_{\overline{c}}(h\overline{u})}{h},\ E_{\overline{c}}(h\overline{u}) \to \overline{0} \text{ as } h \to 0 \\
&\implies \lim_{h\to 0} \frac{f(\overline{c}+h\overline{u}) - f(\overline{c})}{h} = T_{\overline{c}}(\overline{u}) + \lim_{h\to 0} \frac{|h|\|\overline{u}\| E_{\overline{c}}(h\overline{u})}{h} \\
&\implies f'(\overline{c}, \overline{u}) = T_{\overline{c}}(\overline{u})
\end{aligned}
$$

□

Note. *$T_{\overline{c}}$ is linear, however $E_{\overline{c}}$ is not linear. Thus $E_{\overline{c}}(h\overline{u}) \neq hE_{\overline{c}}(\overline{u})$.*

As $h \to 0$, $h\overline{u} \to \overline{0}$ and $E_{\overline{c}}(h\overline{u}) \to \overline{0}$. Since the order of the function $E_{\overline{c}}(h\overline{u})$ is much smaller than that of h, the limit on the right converges to 0.

Theorem 16.4. *If f is differentiable at $\overline{c}$, then f is continuous at $\overline{c}$.*

Proof. Let $\overline{v} \neq 0$, then

$$\begin{aligned} \overline{v} &= v_1\overline{u_1} + v_2\overline{u_2} + \cdots + v_n\overline{u_n}, \\ \overline{v} \to \overline{0} &\implies \forall j,\ v_j \to 0 \\ T \text{ is linear} &\implies T_{\overline{c}}(\overline{v}) = v_1 T_{\overline{c}}(\overline{u_1}) + v_2 T_{\overline{c}}(\overline{u_2}) + \cdots + v_n T_{\overline{c}}(\overline{u_n}) \\ &\text{Thus, } T_{\overline{c}}(\overline{v}) \to \overline{0} \text{ as } \overline{v} \to 0 \end{aligned}$$

Since f differentiable at $\overline{c}$, there exists linear function $T_{\overline{c}}$ such that

$$\begin{aligned} f(\overline{c}+\overline{v}) &= f(\overline{c}) + T_{\overline{c}}(\overline{v}) + \|v\|E_{\overline{c}}(\overline{v}) \\ &\implies \lim_{\overline{v}\to\overline{0}} f(\overline{c}+\overline{v}) = f(\overline{c}) + \lim_{\overline{v}\to\overline{0}} T_{\overline{c}}(\overline{v}) + \lim_{\overline{v}\to\overline{0}} \|v\|E_{\overline{c}}(\overline{v}) \\ &\implies \lim_{\overline{v}\to\overline{0}} f(\overline{c}+\overline{v}) = f(\overline{c}) \end{aligned}$$

□

Theorem 16.5. *Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}^m$ be differentiable at an interior point $\overline{c}$ of S, where $S \subseteq \mathbb{R}^n$. If $\overline{v} = v_1\overline{u_1} + v_2\overline{u_2} + \cdots + v_n\overline{u_n}$, then*

$$f'(\overline{c})(\overline{v}) = \sum_{k=1}^{n} v_k D_k f(\overline{c})$$

In particular, if f is real-valued $(m = 1)$ we have, $f'(\overline{c})(\overline{v}) = \nabla f(\overline{c}).\overline{v}$

Proof. Suppose $f : S \to \mathbb{R}^m$ is differentiable at $\overline{c}$, then there exists a linear function $f'(\overline{c}) : S \to \mathbb{R}^m$ such that $f(\overline{c}+\overline{v}) = f(\overline{c}) + f'(\overline{c})(\overline{v}) + \|\overline{v}\|E_{\overline{c}}(\overline{c})$ where $E_{\overline{c}} \to \overline{0}$ as $\overline{v} \to \overline{0}$.

$$\begin{aligned} f'(\overline{c})(\overline{v}) &= f'(\overline{c})\left(\sum_{k=1}^{n} v_k\overline{u_k}\right) \\ &= \sum_{k=1}^{n} v_k f'(\overline{c})(\overline{u_k}),\ \text{since } f'(\overline{c}) \text{ is linear} \\ &= \sum_{k=1}^{n} v_k D_k f(\overline{c}),\ \text{since } f'(\overline{c})(\overline{u_k}) = f'(\overline{c}, \overline{u_k}) = D_k f(\overline{c}) \end{aligned}$$

Let $m = 1$, then $f : S \to \mathbb{R}$

$$\begin{aligned} f'(\overline{c})(\overline{v}) &= \sum_{k=1}^{n} v_k D_k f(\overline{c}) = \nabla f(\overline{c}).\overline{v} \\ &\text{since } \nabla f(\overline{c}) = (D_1 f(\overline{c}),\ D_2 f(\overline{c}),\ \cdots,\ D_n f(\overline{c})) \end{aligned}$$

□

Remark. *Let $f : S \to \mathbb{R}$, then $f(\overline{c}+\overline{v}) = f(\overline{c}) + \nabla f(\overline{c}).\overline{v} + o(\|\overline{v}\|)$ as $\overline{v} \to \overline{0}$.*

Remark (Complex-valued Functions)**.** *Seminar - Rohitha*

16.3 Matrix of Linear Function

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear function. Let $\{\overline{u_1},\ \overline{u_2},\ \cdots,\ \overline{u_n}\}$ be standard basis for $\mathbb{R}^n$ and $\{\overline{e_1},\ \overline{e_2},\ \cdots,\ \overline{e_m}\}$ be standard basis for $\mathbb{R}^m$. Let $\overline{v} \in \mathbb{R}^n$, then $T(\overline{v}) = \sum_{k=1}^{n} v_k T(\overline{u_k})$ and

$$T(\overline{v}) = \begin{bmatrix} T(\overline{u_1}) & T(\overline{u_2}) & \cdots & T(\overline{u_n}) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} t_{11}\overline{e_1} + \cdots + t_{1m}\overline{e_m} & t_{21}\overline{e_1} + \cdots + t_{2m}\overline{e_m} & \cdots & t_{n1}\overline{e_n} + \cdots + t_{nm}\overline{e_m} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

$$= \begin{bmatrix} \overline{e_1} & \overline{e_2} & \cdots & \overline{e_m} \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1m} \\ t_{21} & t_{22} & \cdots & t_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nm} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{bmatrix}$$

$$T(\overline{v}) = T\left(\sum_{k=1}^{n} v_k \overline{u_k}\right) = \sum_{k=1}^{n} v_k T(\overline{u_k}) = \sum_{k=1}^{n} v_n \sum_{j=1}^{m} t_{kj}\overline{e_j}$$

Thus matrix of T is given by, $m(T) = (t_{ik})$ where $T(\overline{u_k}) = \sum_{k=1}^{n} t_{ik}\overline{e_i}$.

Remark. *Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (2x + y, y - z)$. Then*

$$m(T) = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & -1 \end{bmatrix},\ T(1, 2, 3) = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \overline{e_1} \\ \overline{e_2} \end{bmatrix} = 4\overline{e_1} - \overline{e_2} = (4, -1)$$

Part IV

ME800402 Algorithmic Graph Theory

Chapter 17

An Introduction to Graphs

Chapter 18

An Introduction to Algorithms

Chapter 19

Trees

Chapter 20

Paths and Distance in Graphs

Chapter 21

Networks

21.1 An Introduction to Networks

Definitions 21.1. *A **network** N is a digraph D with two special vertices source s and sink t together with a capacity function $c : E(D) \to \mathbb{Z}$ such that for every arc $a = (u, v)$ of the digraph, $c(u, v)$ is non-negative.*

Remark. *Mathematical Modeling using Network,*

1. *There is no restriction on indegree/outdegree of source/sink vertices of the digraph D of a network N.*
2. *Applications of Network : Transportation problem.*

$c(u, v)$ is the capacity of the arc (u, v) of D

$N^+(x) = \{y \in V(D) : (x, y) \in E(D)\}$ is the out-neighbourhood of x.

$N^-(x) = \{y \in V(D) : (y, x) \in E(D)\}$ is the in-neighbourhood of x.

Definitions 21.2. *A **flow f in a network N** is function $f : E(D) \to \mathbb{Z}$ such that 1. each edge satisfies capacity constraint and 2. each vertex except source and sink satisfies conservation equation.*

capacity constraint

$$0 \leq f(a) \leq c(a) \text{ for every arc } a \in V(D) \tag{21.1}$$

conservation equation

$$\sum_{y \in N^+(x)} f(x, y) = \sum_{y \in N^-(x)} f(y, x), \ \forall \text{vertex } x \in V(D) - \{s, t\} \tag{21.2}$$

net flow out of x

$$\sum_{y \in N^+(x)} f(x, y) - \sum_{y \in N^-(x)} f(y, x)$$

net flow into x

$$\sum_{y \in N^-(x)} f(y, x) - \sum_{y \in N^+(x)} f(x, y)$$

Definitions 21.3. *The **flow f in a network N** is the net flow out of source s.*

Remark. *1. net flow out of/into $x \in V(D) - \{s, t\}$ is zero.*

2. Without loss of generality[1], underlying digraph is always assymetric.

$(X, Y) = \{(x, y) \in E(D) : x \in X,\ y \in Y\}$.
Let X, Y be non-empty subsets of $V(D)$ such that X, Y are disjoint. Then (X, Y) is the set of all arcs from X to Y.

flow from X to Y is the sum of flow on each arc in (X, Y)

$$f(X, Y) = \sum_{(x,y)\in(X,Y)} f(x, y) \tag{21.3}$$

capacity of the partition (X, Y) is the total capacity of arcs in (X, Y)

$$c(X, Y) = \sum_{(x,y)\in(X,Y)} c(x, y) \tag{21.4}$$

cut Let $P \subset V(D)$ such that $s \in P$ and $t \notin P$ and $\bar{P} = V(D) - P$, then $(P, \bar{P})$ is a cut.

flow from P to $\bar{P}$ is the sum of flow on each arc in $(P, \bar{P})$.

$$f(P, \bar{P}) = \sum_{(x,y)\in(P,\bar{P})} f(x, y) \tag{21.5}$$

flow from $\bar{P}$ to P is the sum of flow on each arc in $(\bar{P}, P)$

$$f(\bar{P}, P) = \sum_{(x,y)\in(\bar{P},P)} f(x, y) \tag{21.6}$$

capacity of the cut $(P, \bar{P})$ is the total capacity of the arcs in $(P, \bar{P})$

$$c(P, \bar{P}) = \sum_{(x,y)\in(P,\bar{P})} c(x, y) \tag{21.7}$$

Theorem 21.4. *For any cut $(P, \bar{P})$, the flow in N is $f(N) = f(P, \bar{P}) - f(\bar{P}, P)$.*

Synopsis. *The net flow out of source s is the flow $f(N)$ in the network N. Let $(P, \bar{P})$ be a cut of N, then $s \in P$ and $t \notin P$. Suppose $P = \{s\}$, then the theorem is true. Suppose P is not singleton, then for each vertex $x \in P$, $x \neq s$, the net flow out of x is zero by flow conservation equation. And flow between vertices in P cancels out each other. Thus adding net flow out of each vertex in P, will be same as the net flow out of source which is the flow in the network, $f(N)$.*

[1] If underlying digraph of a network is symmetric, then by replacing an arc (u, v) with a new vertex w and two arcs $(u, w), (w, v)$ gives an assymetric digraph.[Gray Chartrand,]pp.131

Proof.

$$\text{Flow, } f = \sum_{y \in N^+(s)} f(s,y) - \sum_{y \in N^-(s)} f(y,s) \tag{21.8}$$

By conservation equation, we have $\forall x \in P,\ x \neq s$,

$$\sum_{y \in N^+(x)} f(x,y) - \sum_{y \in N^-(x)} f(y,x) = 0 \tag{21.9}$$

By above equations,

$$\begin{aligned} \text{Flow, } f &= \sum_{x \in P} \sum_{y \in N^+(x)} f(x,y) - \sum_{x \in P} \sum_{y \in N^-(x)} f(y,x) \\ &= \sum_{(x,y) \in (P,\bar{P})} f(x,y) - \sum_{(y,x) \in (\bar{P},P)} f(y,x) \end{aligned} \tag{21.10}$$

□

Corollary 21.4.1. *Flow cannot exceed the capacity of any cut $(P, \bar{P})$. Further, $f(N) \leq \min c(P, \bar{P})$.*

Synopsis. *Let $(P, \bar{P})$ be a cut in network N, then by theorem the flow $f(N) =$ flow from P to $\bar{P}$ - flow from $\bar{P}$ to P. Since the flow from $\bar{P}$ to P is non-negative, $f(N) \leq$ flow from P to $\bar{P}$. Clearly, $f(x,y) \leq c(x,y)$ by the capacity constaint. Thus $f(N) \leq f(P,\bar{P}) \leq c(P,\bar{P}) \leq \min c(P,\bar{P})$.*

Proof.

$$\begin{aligned} f(N) &= \sum_{(x,y) \in (P,\bar{P})} f(x,y) - \sum_{(y,x) \in (\bar{P},P)} f(y,x) \\ &\leq \sum_{(x,y) \in (P,\bar{P})} f(x,y) = f(P,\bar{P}) \\ &\leq \sum_{(x,y) \in (P,\bar{P})} c(x,y) = c(P,\bar{P}), \quad \because \forall x,y \in V(D),\ f(x,y) \leq c(x,y) \\ &\leq \min c(P,\bar{P}) \end{aligned}$$

□

Corollary 21.4.2. *In a network N flow is the net flow into the sink of N.*

Synopsis. *Let $\bar{P} = \{t\}$, then by theorem $f(N)$ is the net flow into the sink.*

Proof. Suppose $P = V(D) - \{t\}$. Then by theorem, we have

$$\begin{aligned} f(N) &= \sum_{(x,y) \in (P,\bar{P})} f(x,y) - \sum_{(y,x) \in (\bar{P},P)} f(y,x) \\ &= \sum_{x \in N^-(t)} f(x,t) - \sum_{x \in N^+(t)} f(t,x) \end{aligned}$$

□

Remark. *Exercise 5.1*

4. *Let N be a network with underlying digraph D which has a vertex $v \in V(D) - \{s,t\}$ with zero indegree. Clearly the flow into v is zero. Thus flow out of v is also zero by flow conservation equation. Let N' be the network obtained from N by deleting the vertex v. Then $f(N) = f(N')$.*

21.2 The Max-Flow Min-Cut Theorem

maximum flow A flow f in network N is maximum flow in N, if $f(N) \geq f'(N)$ for each flow f' in N.

minimum cut A cut $(P, \bar{P})$ in network N is minimum cut of N, if $c(P, \bar{P}) \leq c(X, \bar{X})$ for each cut $(X, \bar{X})$ in N.

f-unsaturated Let f be a flow in network N with underlying digraph D, and $Q = u_0, a_1, u_1, a_2, \cdots, u_{n-1}, a_n, u_n$ be a semipath in D such that every forward arc $a_i = (u_{i-1}, u_i)$ has flow not upto its capacity, $f(a_i) < c(a_i)$ and every reverse arc $a_i = (u_i, u_{i-1})$ has some positive flow in it, $f(a_i) > 0$

f-augmenting semipath Let f be a flow in a network N with underlying digraph D. Suppose semipath $Q = s, a_1, u_1, a_2, \cdots, u_{n-1}, a_n, t$ (from source to sink) is f-unsaturated, then Q is an f-augmenting semipath.

Theorem 21.5. *Let f be a flow in a network N with underlying digraph D. The flow f is maximum in N iff there is no f-augmenting semipath in D.*

Synopsis. *Suppose Q is an f-augmenting semipath in D, then there exists a flow f^* in N such that $f(N)+\Delta = f^*(N)$. Therefore, f is not a maximum flow in N. Suppose there is no f-augmenting semipath in D, then there exists a cut $(P, \bar{P})$ such that $f(a) = c(a)\ \forall a \in (P, \bar{P})$ and $f(a) = 0\ \forall a \in (\bar{P}, P)$. Suppose f^* in a maximum flow in N, then $f(N) \leq f^*(N) \leq c(P, \bar{P}) = f(N)$.*

Proof. Let f be a flow in a network N with underlying digraph D and $Q = s, a_1, u_1, a_2, u_2, \cdots, u_{n-1}, a_n, t$ be an f-augmenting semipath in D.

$$\text{define } \Delta_i = \begin{cases} c(a_i) - f(a_i) \text{ for every forward arc } a_i \in Q, \\ f(a_i) \text{ for every reverse arc } a_i \in Q, \end{cases}$$

Define $\Delta = \min\{\Delta_i\}$. Also define $f^* : E(D) \to \mathbb{Z}$ such that

$$f^*(a_i) = \begin{cases} f(a) + \Delta, \text{ for every forward arc } a_i \in Q, \\ f(a) - \Delta, \text{ for every reverse arc } a_i \in Q, \\ f(a_i), \text{ for every arc of } D \text{ which are not in } Q. \end{cases}$$

Since Q is an f-augmenting semipath in D, $\Delta > 0$ and $f(N) + \Delta = f^*(N)$.

Clearly $f(N) < f^*(N)$, and it is enough to show that f^* is a flow in N. f^* is a flow if it satisfies 1. capacity constraint and 2. conservation equation. For any arc $a_i \notin Q$, $f^*(a_i) = f(a_i) \leq c(a_i)$. Suppose $a_i \in Q$. If $a_i = (u_{i-1}, u_i)$, a_i is a forward arc and we have $f^*(a_i) = f(a_i)+\Delta \leq f(a_i)+\Delta_i = f(a_i) + c(a_i) - f(a_i) = c(a_i)$. If $a_i = (u_i, u_{i-1})$, then a_i is a reverse arc and we have $f^*(a_i) = \Delta \leq \min\{\Delta_i\} = \Delta_i = c(a_i)$. Thus f^* satisfies capactity constraint on every arc of D.

Let $x \in V(D) - \{s,t\}$. Suppose $x \notin Q$,

$$\begin{aligned}\text{Net flow out of } x &= \sum_{y\in N^+(x)} f^*(x,y) - \sum_{y\in N^-(x)} f^*(y,x)\\ &= \sum_{y\in N^+(x)} f(x,y) - \sum_{y\in N^-(x)} f(y,x)\\ &= 0\end{aligned}$$

Suppose $x = u_i \in Q$, then Q has two arc having vertex x say, a_{i-1}, and a_i. There are four possibilities for these two arcs,

1. Both $a_{i-1},\ a_i$ are forward arcs.
2. Arc a_{i-1} is forward, but arc a_i is reverse.
3. Arc a_{i-1} is reverse, but arc a_i is forward.
4. Both $a_{i-1},\ a_i$ are reverse arcs.

Case 1 $a_{i-1} = (u_{i-1}, u_i)$ and $a_i = (u_i, u_{i+1})$.

$$\begin{aligned}\text{Net flow out of } x &= \sum_{y\in N^+(x)} f^*(x,y) - \sum_{y\in N^-(x)} f^*(y,x)\\ &= \sum_{\substack{y\in N^+(x)\\ y\neq u_{i+1}}} f^*(x,y) + f^*(u_i,u_{i+1}) - \left(\sum_{\substack{y\in N^-(x)\\ y\neq u_{i-1}}} f^*(y,x) + f^*(u_{i-1},u_i)\right)\\ &= \sum_{\substack{y\in N^+(x)\\ y\neq u_{i+1}}} f(x,y) + f(u_i,u_{i+1}) + \Delta - \left(\sum_{\substack{y\in N^-(x)\\ y\neq u_{i-1}}} f(y,x) + f(u_{i-1},u_i)\right) - \Delta\\ &= \sum_{y\in N^+(x)} f(x,y) - \sum_{y\in N^-(x)} f(y,x)\\ &= 0\end{aligned}$$

Case 2 $a_{i-1} = (u_{i-1}, u_i)$ and $a_i = (u_{i+1}, u_i)$.

$$\begin{aligned}\text{Net flow out of } x &= \sum_{y\in N^+(x)} f^*(x,y) - \sum_{y\in N^-(x)} f^*(y,x)\\ &= \sum_{\substack{y\in N^+(x)\\ y\neq u_{i+1},u_{i-1}}} f^*(x,y) + f^*(u_i,u_{i+1}) + f^*(u_i,u_{i-1}) - \sum_{y\in N^-(x)} f^*(y,x)\\ &= \sum_{\substack{y\in N^+(x)\\ y\neq u_{i+1},u_{i-1}}} f(x,y) + f(u_i,u_{i+1}) + \Delta + f(u_i,u_{i-1}) - \Delta - \sum_{y\in N^-(x)} f(y,x)\\ &= \sum_{y\in N^+(x)} f(x,y) - \sum_{y\in N^-(x)} f(y,x)\\ &= 0\end{aligned}$$

Case 3 $a_{i-1} = (u_i, u_{i-1})$ and $a_i = (u_i, u_{i+1})$.

$$\begin{aligned}
\text{Net flow out of } x &= \sum_{y \in N^+(x)} f^*(x,y) - \sum_{y \in N^-(x)} f^*(y,x) \\
&= \sum_{y \in N^+(x)} f^*(x,y) - \left(\sum_{\substack{y \in N^-(x) \\ y \neq u_{i-1}, u_{i+1}}} f^*(y,x) + f^*(u_{i-1}, u_i) + f^*(u_{i+1}, u_i) \right) \\
&= \sum_{y \in N^+(x)} f^*(x,y) - \left(\sum_{\substack{y \in N^-(x) \\ y \neq u_{i-1}, u_{i+1}}} f(y,x) + f(u_{i-1}, u_i) + \Delta + f(u_{i+1}, u_i) - \Delta \right) \\
&= \sum_{y \in N^+(x)} f(x,y) - \sum_{y \in N^-(x)} f(y,x) \\
&= 0
\end{aligned}$$

Case 4 $a_{i-1} = (u_i, u_{i-1})$ and $a_i = (u_{i+1}, u_i)$.

$$\begin{aligned}
\text{Net flow out of } x &= \sum_{y \in N^+(x)} f^*(x,y) - \sum_{y \in N^-(x)} f^*(y,x) \\
&= \sum_{\substack{y \in N^+(x) \\ y \neq u_{i-1}}} f^*(x,y) + f^*(u_i, u_{i-1}) - \left(\sum_{\substack{y \in N^-(x) \\ y \neq u_{i+1}}} f^*(y,x) + f^*(u_{i+1}, u_i) \right) \\
&= \sum_{\substack{y \in N^+(x) \\ y \neq u_{i-1}}} f(x,y) + f(u_i, u_{i-1}) - \Delta - \left(\sum_{\substack{y \in N^-(x) \\ y \neq u_{i+1}}} f(y,x) + f(u_{i+1}, u_i) \right) + \Delta \\
&= \sum_{y \in N^+(x)} f(x,y) - \sum_{y \in N^-(x)} f(y,x) \\
&= 0
\end{aligned}$$

Therefore, f^* is a flow on N. We have $f(N) < f^*(N)$. Thus f is not maximum flow in N due to the existence of an f-augmenting semipath in D.

Conversely, assume that there is no f-augmenting semipath in D. Now, we construct a cut $(P, \bar{P})$ of N. Let P be the set of all vertices $x \in V(D)$ such that there is an f-unsaturated $s - x$ semipath in D. Trivially, $s \in P$. And $t \notin P$ since there are no f-augmenting semipath in D.[2] Clearly, $(P, \bar{P})$ is a cut of the network N.

We claim that $c(P, \bar{P}) = f(N)$. Suppose there is a forward arc $(x, y) \in (P, \bar{P})$, then flow in it is saturated. If $f(x, y) < c(x, y)$, then there is an f-unsaturated $s - y$ semipath in D. ie, $s - x$ semipath + arc (x, y). Thus every forward arc $(x, y) \in (P, \bar{P})$ is saturated. Suppose there is a reverse arc $(y, x) \in$

[2] An f-augmenting semipath is an f-unsaturated $s - t$ semipath in D.

$(\bar{P}, P)$, then there is no flow in it(saturated reversed arc). If $f(y,x) > 0$, then there is an f-unsaturated $s-y$ semipath in D. ie, $s-x$ semipath + arc (y,x). Thus every reverse arc $(y,x) \in (\bar{P}, P)$ is saturated. And we have,

$$\sum_{(x,y)\in(P,\bar{P})} f(x,y) = \sum_{(x,y)\in(P,\bar{P})} c(x,y)$$

$$\sum_{(y,x)\in(\bar{P},P)} f(y,x) = 0$$

$$\begin{aligned} f(N) &= \sum_{(x,y)\in(P,\bar{P})} f(x,y) - \sum_{(y,x)\in(\bar{P},P)} f(y,x) \\ &= \sum_{(x,y)\in(P,\bar{P})} c(x,y) \\ &= c(P,\bar{P}) \end{aligned}$$

Suppose f^* is maximum flow in network N and $(X, \bar{X})$ is minimum cut of N. Then $f(N) \leq f^*(N)$. Thus we have, $f(N) \leq f^*(N) \leq c(X,\bar{X}) \leq c(P,\bar{P}) = f(N)$. Therefore, $f(N) = f^*(N)$. ie, the flow f is maximum in network N if there are no f-augmenting semipaths in D. □

Theorem 21.6 (maximum-flow, min-cut)**.** *In every network, the value of maximum flow equals capacity of minimum cut.*

Proof. Suppose flow f in network N in maximum, then by previous theorem there is no f-augmenting semipath in D. And $f(N) \leq c(X,\bar{X})$ for any cut $(X,\bar{X})$ in N. We can construct a cut $(P,\bar{P})$ in N such that $f(N) = c(P,\bar{P})$. Let P be the set of all vertices x in D such that there is an f-unsaturated $s-x$ semipath in D. Clearly $s \in P$ and $t \notin P$. Also $f(P,\bar{P}) = c(P,\bar{P})$ and $f(\bar{P},P) = 0$. Then the cut $(P,\bar{P})$ is minimum cut of N. Suppose there is a cut $(X,\bar{X})$ such that $c(X,\bar{X}) < c(P,\bar{P})$. Then $f(N) = f(P,\bar{P}) - f(\bar{P},P) = c(P,\bar{P}) < c(X,\bar{X})$ which is a contradiction. Therefore, the value of maximum flow equals capacity of minimum cut. □

Remark. *Exercise 5.2*

1. *Suppose $(X,\bar{X})$ is a cut of N such that $f(a) = c(a)$, $\forall a \in (X,\bar{X})$ and $f(a) = 0$, $\forall a \in (\bar{X}, X)$. By the definition of cut, $s \in X$ and $t \in \bar{X}$. Thus there is no f-augmenting semipath in D. Suppose there is an f-augmenting semipath Q in D, then there is either (a) a forward arc $(x,y) \in (X,\bar{X})$ such that $f(x,y) < c(x,y)$ or (b) a reverse arc $(y,x) \in (\bar{X}, X)$ such that $f(y,x) > 0$ which is a contradition. Therefore, the flow $f(N)$ is maximum and the given cut $(X,\bar{X})$ is minimum as shown in the proof of the maximum-flow min-cut theorem.*

3. *The algorithm suggested in the hint of this exercise won't work if two subnetworks have a common arc such that the direction of flow in which is not consistent. Suppose, the generalized network is not supposed to have any common arcs. Then construct subnetworks for each pair (s,t) with all those arcs which are on some $s-t$ semipath. Define subnetwork capacity function $c'(a) = c(a)$ for every arc in N'.*

Let N be a generalized network with set of sources S and set of sinks T. A flow in N is maximum if there is not f-augmenting $s-t$ semipath for each pair $(s,t) \in S \times T$.

21.3 A max-flow min-cut algorithm

Theorem 21.7. *Let N be a network with underlying digraph D, source s, sink t, capacity function c and flow f. Let D' be the digraph with same vertex set as D and arc set defined by $E(D') = \{(x,y) : (x,y) \in E(D),\ c(x,y) > f(x,y)$ or $(y,x) \in E(D),\ f(y,x) > 0\}$. ie, D' has only the unsaturated arcs of D. Then D' has an $s-t$ directed path iff D has an f-augmenting semipath. Moreover, shortest $s-t$ path in D' has the same length as shortest f-augmenting semipath in D.*

Synopsis. *Each directed $s-t$ path in D' has respective f-augmenting semipath in D and vice versa. Clearly, they have the same length.*

Proof. Let N be a network with underlying digraph D, capacity c and flow f. Let D' be the digraph with vertex set $V(D') = V(D)$ and arc set $E(D') = \{(x,y) :$ either (x,y) or (y,x) is unsaturated in $N\}$.

Suppose D' has a directed $s-t$ path $Q' : s, u_1, u_2, \cdots, u_{n-1}, t$. Then by the construction of D', for each $u_i \in Q$, there exists an f unsaturated arc a_i in D. ie, either forward arc $a_i = (u_{k-1}, u_k)$ such that $f(u_{k-1}, u_k) < c(u_{k-1}, u_k)$ or reverse arc $a_i = (u_k, u_{k-1})$ such that $f(u_k, u_{k-1}) > 0$. Therefore, we have an $s-t$ semipath $Q : s, a_1, u_1, a_2, \cdots, u_{n-1}, a_n, t$ in D such that Q is an f-augmenting semipath since every arc in Q is f-unsaturated. Clearly, Q, Q' are of the same length.

Conversely, suppose that the digraph D has an f-augmenting semipath $Q : s, a_1, u_1, a_2, \cdots, u_{n-1}, a_n, t$. Then each arc $a_i \in Q$ are f-unsaturated and by the construction of D', there exists a directed $s-t$ path $Q' = s, u_1, u_2, \cdots, u_{n-1}, t$ in D'. And Q, Q' are of the same length.

There is a one-one correspondence betweeen the directed $s-t$ paths in D' and f-augmenting semipaths in D. Clearly, they have the same length. Thus shortest directed $s-t$ path in D and shortest f-augmenting semipath in D' are of the same length. □

saturation arc of N with respect to the flow f is an arc a_j in an f-augmenting semipath Q with $\Delta_j = \Delta$.

augmentation path is an f-augmenting semipath Q in D.

Algorithm 21.8 (max-flow min-cut)**.** *An algorithm to find maximum flow and minimum cut of a network N with underlying digraph D, source s, sink t, capacity function c and initial flow f.*

1. *Construct digraph D' with vertex set $V(D') = V(D)$ and arc set $E(D') = \{(x,y) : (x,y) \in E(D) \& f(x,y) < c(x,y)$ or $(y,x) \in E(D) \& f(y,x) > 0\}$*

2. *Find (shortest) $s-t$ directed path in D' using Moore's breadth first search(BFS) algorithm. If D' doesn't have an $s-t$ path, then proceed to step 5. Otherwise, let $Q' : s, u_1, u_2, \cdots, u_{n-1}, t$ be a (shortest) $s-t$ path in D'.*

3. *Let $Q : s, a_1, u_1, a_2, \cdots, u_{n-1}, a_n, t$ be the respective semipath in D such that $f(a_j) < c(a_j)$ for forward arcs and $f(a_i) > 0$ for reverse arcs. Let $\Delta_j = c(a_j) - f(a_j)$ for forward arcs and $\Delta_j = f(a_j)$ for reverse arcs. And let $\Delta = \min\{\Delta_j\}$. And augment flow f by Δ ie, $f(a_j) \leftarrow f(a_j) + \Delta$ for forward arcs and $f(a_j) \leftarrow f(a_j) - \Delta$ for reverse arcs.*

4. *Goto step 1 (Proceed with new flow f and find whether there are any directed $s-t$ paths in D'. If any, augment the flow along the new augmentation path Q by saturating the flow along the saturation arc.)*

5. *There is no $s-t$ directed path in D'. Thus there is no f-augmenting semipath in D. Therefore the flow f in N is maximum. Let P be the set of all vertices in D' with non-zero breadth first index(bfi) from Moore's BFS algorithm applied in step 2. $(P, \bar{P})$ is minimum cut of N.*

Remark. *Validity of the algorithm is proved in the previous theorem.*

21.5 Connectivity and Edge-Connectivity

edge cutset is the set U subset of $E(G)$ such that $G - U$ is disconnected.

vertex cutset is the set S subset of $V(G)$ such that $G - S$ is disconnected.

edge connectivity $\lambda(G)$ is the minimum cardinality of all edge cutsets of G.

connectivity $\kappa(G)$ is the minimum cardinality of all vertex cutsets of G.

Theorem 21.9. *For every graph G, $\kappa(G) \leq \lambda(G) \leq \delta(G)$*

Proof. Suppose graph G is disconnected then $\kappa(G) = \lambda(G) = \delta(G) = 0$. Let G be a connected graph. Then G has at least one vertex v with degree $\delta(G)$. Therefore $\lambda(G) \leq \delta(G)$ since edges incident with v form an edge cutset of G and $\lambda(G)$ is the cardinality of all edge cutsets.

Let G be a graph with edge connectivity $\lambda(G) = c$. Let U be a edge cutset with cardinality c and let edge $uv \in U$. Construct a set of vertices $S \subset V(G)$ such that (S is of minimal cardinality and) for each edge in U other uv, S has a vertex incident with it. Cardinality of S is atmost $c - 1$, since we can select one vertex each for each edge in U other than u, v. If $G - S$ is a disconnected graph, then $\kappa(G) < \lambda(G)$. Suppose $G - S$ is a connected graph, then delete a non-pendent vertex u or v from $G - S$, say v. Since $G - S$ is a connected graph with a singleton edge cutset, $\{uv\}$. We have a vertex cutset $S \cup \{v\}$ of G. Therefore, $\kappa(G) \leq c = \lambda(G)$. □

Theorem 21.10. *If G is a graph of diameter 2, then $\lambda(G) = \delta(G)$*

n-edge connected G is n-edge connected if $\lambda(G) \geq n$.

n connected G is n-connected if $\kappa(G) \geq n$.

Theorem 21.11. *Let G be a graph of order p and n be an integer such that $1 \leq n \leq p-1$. If $\delta(G) \geq \frac{p+n-2}{2}$, then G is n-connected.*

connection number $c(G)$ is the smallest integer such that $2 \leq c(G) \leq p$ and every subgraph of order n in G is connected.

l-connectivity $\kappa_l(G)$ is minimum number of vertices whose removal will produce a disconnected graph with at least l components or a graph with fewer than l vertices.

(n,l)-connected A graph G is (n,l)-connected if $\kappa_l(G) \geq n$.

Remark. *Exercises 5.5*

1. $\lambda(K_{m,n}) = \kappa(K_{m,n}) = m$

8. $c(K_p) = 2$, $c(K_{m,n}) = n+1$, $c(C_p) = p-1$
   Every two vertices of complete graph of order p are adjacent. For complete bi-partitie graph $K_{m,n}$ such that $1 \leq m \leq n$, there exists a totally disconnected subgraph of order n. Therefore $c(K_{m,n}) \geq n+1$. And with $n+1$ vertices, both partitions have at least two vertices each and therefore the graph is connected and $c(K_{m,n}) \leq n+1$. For cycle C_p, any subgraph is disconnected if two non-adjacent vertices are deleted. Therefore $c(C_p) \geq p-1$. And C_p remains connected even after deletion of any vertex, therefore $c(C_p) \leq p-1$.

9. $$\delta(G) \geq \frac{p+(l-1)(n-2)}{l} \implies \kappa_l(G) \geq n$$

21.6 Menger's Theorem

Theorem 21.12. *For a non-trivial graph G, $\lambda(u,v) = M'(u,v)$ for every pair (u,v) of vertices of G.*

Corollary 21.12.1. *Graph G is n-edge connected iff every two vertices of G are connected by at least n edge disjoint paths.*

Theorem 21.13. *For every pair of non-adjacent vertices u, v in graph G, $\kappa(u,v) = M(u,v)$.*

Corollary 21.13.1. *Graph G is n-connected iff every pair of vertices of G are connected by at least n internally disjoint paths.*

Algorithm 21.14 (connectivity $\kappa(G)$)**.** .

1. *If degree of every vertex is $p-1$, then output $\kappa = p-1$ and stop. Otherwise, continue.*

2. *If G is disconnected, output $\kappa = 0$ and stop. Otherwise, continue.*

3. $\kappa \leftarrow p$

4. $i \leftarrow 0$

5. *If $i \leq \kappa$, then $i \leftarrow i+1$ and continue. Otherwise, output κ and stop.*

6. $j \leftarrow i+1$

7. *(1) If $j = p+1$, then return to step 5. Otherwise continue.*

   (2) If $v_i v_j \notin E(G)$, construct network N with digraph D as follows : for each vertex $v \in V(G)$, there are two vertices $v', v'' \in V(D)$ and an arc $(v', v'') \in E(D)$. And for each edge $uv \in E(G)$, there are two arcs $(u'', v), (v'', u) \in E(D)$. The capacity function is given by, $c(v', v'') = 1$ for every $v \in V(G)$ and $c(a) = \infty$ for every other arc in D. Set source $s = v_i''$ and sink $t = v_j'$ and find maximum flow in N using max-flow min-cut algorithm. Otherwise proceed to step 7d

   (3) If $f(N) < \kappa$, then $\kappa \leftarrow f(N)$. Otherwise, continue.

   (4) $j \leftarrow j+1$ and return to step 7a

Chapter 22

Matchings and Factorizations

22.1 An Introduction to Matching

Marriage Problem Given a collection of men and women, where each womon knows some of the men. Can every women marry a man she knows ?

Assignment Problem Given several job openings and applicants for one or more of these positions. Find an assignment so that maximum positions are filled ?

Optimal Assignment Problem Given several job openings and applicants for one or more of these positions. The benefits of employing these applicants on those positions are also given. Find an assignment of maximum benefit to the company ?

matching in G is a 1-regular[1] subgraph of G.

maximum matching in G is a matching of G with maximum cardinality.

perfect matching in G is a matching of cardinality $p/2$. ie, $p/2$ edges.

maximum weight matching in a weighted graph G is a matching with maximum weight.

Definitions 22.1. *Let M be a matching in a graph G,*

matched edge *is an edge in subgraph M of G.*

unmatched edge *is an edge of G that doesn't belong to M.*

matched vertex *with respect to M is a vertex incident with an edge of M.*

single vertex *is a vertex that is not incident with any edge of M.*

alternating path *in G is a path with edges alternately matched and unmatched.*

[1] A graph G is k-regular, if every vertex of G has degree k.

augmenting path *in G is a non-trivial alternating path with single vertices as end vertices.*

Theorem 22.2. *Let M_1, M_2 be two matchings in G such that there is a spanning subgraph H of G with edges that are either in M_1 or M_2, but not both. Then the components of H are either 1. isolated vertex 2. even cycle with edge alternately from M_1 and M_2 3. a non-trivial path with edges alternately from M_1 and M_2 such that each end vertex is single with respect to either M_1 or M_2, but not both.*

Synopsis. $\Delta(H) \leq 2$ *by Pigeonhole principle. Any component of H is either a path or a cycle. A cycle with edge alternately from M_1 and M_2 is even. If an end vertex of a non-trivial path is matched with respect to M_1(WLOG), then it is there in $M_1 - M_2$ ie, it is not there in M_2. If there is another edge in M_2 incident with it, then it has to be in H and it will cease to be an end vertex of path component. Therefore, it is unmatched with respect to M_2.*

Theorem 22.3. *A matching M in a graph G is maximum iff there is no augmenting path with respect to M in G.*

Synopsis. *If M is maximum matching and P an M-augmenting path. Since both end-vertices are single, length of P is odd. Let M', M'' be edges of P which are in M and not in M respectively. Then $M - M' + M''$ is a matching of cardinality one greater than that of M which is a contradiction since M is maximum. Conversely, suppose M be a matching such that there no M-augmenting paths in G. Let M' be a maximum matching in G. Then a nontrivial path component of the graph induced by $M\Delta M'$ is of even length otherwise both end-vertices are matched with respect to one of the matching M or M' which is a contradiction. Again every cycle components are even. Therefore $|M| = |M'|$, since $M\Delta M'$ doesn't have a nontrivial component of another kind.*

Definitions 22.4. *Let U_1, U_2 be two nonempty, disjoint, subsets of the vertex set of a graph G. Then U_1* ***is matched to*** *U_2 if there exists a matching M in G such that every edge in M incident with a vertex in U_1 and a vertex in U_2. And every vertex of U_1 (or U_2) is incident with some edge in M. Suppose M^* be a matching such that $M \subset M^*$, then U_1* ***is matched under*** *M^** ***to*** *U_2.*

Definitions 22.5. *Let U be a nonempty set of vertices of a graph G. U is* ***nondeficient***,[2] *if $|N(S)| \geq |S|$ for every nonempty subset S of U.*

Theorem 22.6. *Let G be a bipartite graph with partite sets V_1, V_2. The set V_1 can be matched to a subset of V_2 iff V_1 is nondeficient.*

Corollary 22.6.1. *Every r-regular bipartite multigraph has a perfect matching.*

Theorem 22.7. *A collection $S_1, S_2, \cdots, S_n$ of finite non-empty sets has a system of distinct representatives iff for each k, $0 \leq k \leq n$, the union of any k of these sets contains at least k elements.*

Remark (Hall's Marriage Theorem)**.** *Suppose there are n women. Then every women can marry a man she knows iff each subset of k women $(1 \leq k \leq n)$ colletively knows atleast k men.*

Remark. *Let W be a set of n women. Then there are $2^n - 1$ nonempty subset for W. Thus, Hall's Marriage Theorem suggests that we ensure $|N(S)| \geq |S|$ for every nonempty subset S of W. This method has complexity $O(2^n)$.*

[2] N(S) is the neighbourhood set of all vertices adjacent to some vertex in S

22.2 Maximum Matching in Bipartite Graphs

Definitions 22.8. *Let M be a matching in a graph G and P is an augmenting path with respect to M. Let M' be set of edge in P and M. And M'' be the set of edges in P and not in M. Then $M_1 = (M - M') \cup M''$ is the* ***matching obtained by augmenting M along path P****.*

Remark. $|M_1| = |M| + 1$

Theorem 22.9. *Let M be a a matching of a graph G that is not maximum, and let v be a single vertex with respect to M. Let M_1 denote the mathing obtained by augmenting M along some augmenting path. If G contains an augmenting path with respect to M_1 that has v and an end-vertex, then G contains an augmenting path with respect to M that has v as an end-vertex*

Corollary 22.9.1. *Let M be a matching of a graph G. Suppose that $M = M_1, M_2, \cdots, M_k$ is a finite sequence of matchings of G such that M_i $(2 \leq i \leq k)$ is obtained by augmenting M_{i-1} along some augmenting path. Suppose v is a single vertex with respect to M for which there exists no augmenting path starting at v. Then G does not contain an augmenting path with respect to M_i $(2 \leq i \leq k)$ that has v as an end-vertex.*

Definitions 22.10. *An* ***alternating tree*** *with respect to a matching M is a tree such that every path from it's root are alternating path with respect to M.*

Algorithm 22.11 (Maximum Matching Algorithm for Bipartite Graphs)**.** .

1. *$i \leftarrow 1$ and $M \leftarrow M_1$*
2. *If $i < p$, then continue; otherwise stop.*
3. *If v_i is matched, then $i \leftarrow i + 1$ and return to Step 2; otherwise, $v \leftarrow v_i$ and Q is initialized to contain v only.*
4. *(1) For $j = 1, 2, \cdots, p$ and $j \neq i$, let $TREE(v_j) \leftarrow F$. Also, $TREE(v_j) \leftarrow T$.*
   (2) If $Q = \phi$, then $i \leftarrow i + 1$ and return to Step 2; otherwise, delete a vertex x from Q and continue.
   (3) (1) Suppose that $N(x) = \{y_1, y_2, \cdots, y_k\}$. Let $j \leftarrow 1$.
   (2) If $j \leq k$, then $y \leftarrow y_j$; otherwise return to Step 4.2
   (3) If $TREE(y) = T$, then $j \leftarrow j + 1$ and return to Step 4.3.2. Otherwise, continue.
   (4) If y is incident with a matching edge yz, then $TREE(y) \leftarrow T$, $TREE(z) \leftarrow T$, $PARENT(y) \leftarrow x$, $PARENT(z) \leftarrow y$ and add z to Q, $j \leftarrow j + 1$ and return to Step 4.3.2. Otherwise, y is a single vertex and continue.
   (5) Use $PARENT$ to determine the alternating $v - x$ path P' in the alternating tree. Let P be the augmenting path obtained from P' by adding the path x, y. Proceed to Step 5
5. *Augment M along P to obtain a new matching M'. Let $M \leftarrow M'$, $i \leftarrow i + 1$, and return to Step 2.*

Definitions 22.12. *Let G be a weighted complete bipartite graph with partite sets V_1 and V_2. A **feasible vertex labeling** is a real function $l : V(G) \to \mathbb{R}$ on vertex set of G such that $l(v) + l(u) \geq w(vu)$ where $v \in V_1$ and $u \in V_2$.*

Definitions 22.13. *Consider the function $l : V(G) \to \mathbb{R}$ such that $\forall v \in V_1,\ l(v) = \max\{w(vu) : u \in V_2\}$ and $\forall u \in V_2,\ l(u) = 0$. Then l is a feasible vertex labeling on $V(G)$. And,*

E_l is the set of all edge of the weighted complete bipartite graph G such that $l(v) + l(u) = w(vu)$.

H_l is the spanning subgraph of G induced by the edge set E_l.

Theorem 22.14. *Let l be a feasible vertex labeling of a weighted complete bipartite graph G. If H_l contains a perfect matching M', then M' is a maximum weight matching of G.*

Algorithm 22.15 (Kuhn-Munkres)**.** .

1. *(1) For each $v \in V_1$, let $l(v) \leftarrow \max\{w(uv) : v \in V_2\}$.*
   (2) For each $u \in V_2$, let $l(u) \leftarrow 0$.
   (3) Let H_l be the spanning subgraph of G with edge set E_l.
   (4) Let G_l be the underlying graph of H_l.
2. *Apply Algorithm 22.11 to determine a maximum matching M in G_l.*
3. *(1) If every vertex v of V_1 is matching with respect to M, output M and stop. Otherwise, continue.*
   (2) Let x be the first single vertex of V_1.
   (3) Construct an alternating tree with respect to M that is rooted at x. If an augmenting path P is discovered, then augmenting M along P and return to Step 3.1. Otherwise, let T be the alternating tree with respect to M and rooted at x that cannot be expanded further in G_l.
4. *Compute $m_l \leftarrow \min\{l(v) + l(u) - w(vu) : v \in V_1 \cap V(T), u \in V_2 - V(T)\}$. Let*
$$l'(v) = \begin{cases} l(v) - m_l \text{ for } v \in V_1 \cap V(T) \\ l(v) + m_l \text{ for } v \in V_2 \cap V(T) \\ l(v) \text{ otherwise} \end{cases}$$
5. *Let $l \leftarrow l'$, construct G_l and return to Step 3.3.*

22.4 Factorizations

Definitions 22.16. *A **factor** of a graph G is a spanning[3] subgraph of G.*

Definitions 22.17. *Let $G_1, G_2, \cdots, G_n$ be edge-disjoint factors of G such that $E(G) = \cup_{i=1}^{n} E(G_i)$. Then G is **factorable** and $G = G_1 \oplus G_2 \oplus \cdots \oplus G_n$.*

[3]Spanning subgraph of a graph G has every vertex of G

Definitions 22.18. *An r-regular factor of G is an **r-factor** of G.*

Definitions 22.19. *If G has a factorisation to r-factors, then G is **r-factorable**.*

Remark. *$K_{3,3}$ is 1-factorable. K_5 is 2-factorable.*

Definitions 22.20. *An **odd component of G** is a component of G with odd number of vertices. And an **even component of G** is a component of G of with even number of vertices.*

Theorem 22.21 (Tutte)**.** *A nontrivial graph G has a 1-factor iff for every proper subset S of $V(G)$, the number of odd components of $G - S$ does not exceed $|S|$.*

Remark. *There exist cubic graphs that doesn't have a 1-factor.*

Theorem 22.22 (Petersen)**.** *Every bridgeless cubic graph contains a 1-factor.*

Remark. *Every brideless cubic graphs has a 1-factor. Let G be a bridgeless cubic graph. Consider every pair of factors G_1, G_2 such that $G = G_1 \oplus G_2$ where G_1 is a 1-factor and G_2 is a 2-factor. G is not 1-factorable only if every such G_2 doesn't have a 1-factor.*

Theorem 22.23. *Petersen graph is not 1-factorable.*

Theorem 22.24. *Every r-regular bipartite multigraph ($r \geq 1$) is 1-factorable.*

Remark (Application of 1-factorisation)**.** *For even number p, a 1-factorisation of K_p corresponds to the schedule of a round of the round robin tournament among p teams. If p is odd, consider K_{p+1} where v_{p+1} is an imaginary team called bye team. A game with bye team is a bye.*

Definitions 22.25. *A **hamiltonian cycle** is a spanning cycle. And, **Hamiltonian graph** is a graph containing a hamiltonian cycle.*

Theorem 22.26. *Complete graph K_{2n+1} can be factored into n hamiltonian cycles.*

Remark. *For $n = 3$, K_7 can be factored into three hamiltonian cycles.*

Theorem 22.27. *Let $0 \leq r < p$. Then there exists an r-regular graph of order p iff pr is even.*

Definitions 22.28. *Let $\{E_1, E_2, \cdots, E_n\}$ be partition of $E(G)$. And let H_i be subgraph of G induced by the edge set E_i. A **decomposition** of a graph G is a colletion of these subgraphs $H_1, H_2, \cdots, H_n$. And $G = H_1 \oplus H_2 \oplus \cdots \oplus H_n$.*

Definitions 22.29. *Let $G = H_1 \oplus H_2 \oplus \cdots \oplus H_n$ be a decomposition of G such that $H \cong H_i$. Then G is H-decomposible.*

Remark. *$K_{3,3}$ is $3K_2$-decomposible. K_5 is C_5-decomposible. K_{2n} is nK_2-decomposible. K_{2n+1} is C_{2n+1}-decomposible.*
Every graph is K_2-decomposible. Every complete bipartite graph $K_{m,n}$ is $K_{1,m}$-decomposible and $K_{1,n}$-decomposible.

22.5 Block Designs

Definitions 22.30. *A block design on a set V is a collection of k-element subsets of V such that each element of V appears exactly in r subsets.*

variety The elements of V are called varieties.

block k-element subsets of V are called blocks.

balanced design If each variety appears in exactly r blocks and each pair of varieties appears in exactly λ blocks.

incomplete design If blocks are proper subsets of V. ie, $k < v$.

Definitions 22.31. *A balanced incomplete block design of v varities in b blocks of cardinality k such that each variety appears in exactly r blocks and each pair of varities appears in exactly λ blocks is a (b, v, r, k, λ)-design.*

Theorem 22.32. $bk = vr$

Theorem 22.33. $\lambda(v-1) = r(k-1)$

Corollary 22.33.1. $\lambda < r$

Theorem 22.34 (Fisher's Inequality)**.** $b \geq v$

symmetric design If $b = v$

Theorem 22.35. *In a symmetric (b, v, r, k, λ)-design with even v, $r - \lambda$ is a perfect square.*

steiner triple system (b, v, r, k, λ)-design with $k = 3$, $\lambda = 1$.

Remark. *(b, v, r, k, λ)-designs are incomplete. However, a complete block design (ie, $v = 3$) is also included as a Steiner triple system.*

Theorem 22.36. *Steiner triple system with v varieties exists iff $v = 6n + 1$ or $v = 6n + 3$ or $v = 3$.*

Definitions 22.37 (Kirkman's Schoolgirls Problem)**.** *A class of 15 girls. Parade 15 girls in five rows (3 girls in a row). Is it possible to plan 7 days parade so that two girls are together in a row exactly once ?*

kirkman triple system Steiner triple system with $v = 6n + 3$.

Remark. *It is proved that kirkman triple system exists with $v = 6n + 3$ for every $n \geq 0$.*

Theorem 22.38. *The code consisting of the rows of the incidence matrix of a (b, v, r, k, λ)-design ($b = v$, $r = k$) is t-error correcting, where $t = k - \lambda - 1$.*

Bibliography

[Apostol, 1973] Apostol, T. (1973). *Mathematical Analysis, 2nd edition*. Narosa Publishing House.

[Gray Chartrand,] Gray Chartrand, O. O. (?). *Applied and Algorithmic Graph Theory*. Tata McGraw Hill Company.

[Joshi, 1983] Joshi, K. D. (1983). *Introduction to General Topology*. Wiley Eastern Ltd.

[Kiusalaas, 2013] Kiusalaas, J. (2013). *Numerical Methods in Engineering with Python3*. Cambridge University Press.

[Munkres, 2003] Munkres, J. R. (2003). *Topology, 2nd edition*. Pearson.