7 Matroids and the Greedy Algorithm

7.1 Matroid definitions

- 1935, matroid theory founded by H. Whitney;
- 1965, J. Edmonds pointed out the significance of matroid theory to combinatorial optimization (CO).

Importance: 1) Many CO problems can be formulated as matroid problems, and solved by the same algorithm;

2) We can detect the insight of the CO problems;

3) A special tool for CO.

Definition. Suppose we have a finite ground set S, $|S| < \infty$, and a collection, Ξ, of subsets of S. Then $H := (S, \Xi)$ is said to be an *independent system* if the empty set is in Ξ and Ξ is closed under inclusion; that is

$$i)\ \emptyset \in \Xi;$$

$$ii)\ X \subseteq Y \in \Xi \Longrightarrow X \in \Xi.$$

Elements in Ξ are called *independent sets*, and subsets of S not in Ξ are called *dependent sets*.

Example: Matching system. $G = (V, E)$,

$$\Xi = \{\text{all matchings in G}\}.$$

Definition. If $H = (S, \Xi)$ is an independent system such that

$$X, Y \in \Xi,\ |X| = |Y| + 1 \implies \text{there exists } e \in X \backslash Y \text{ such that } Y + e \in \Xi,$$

then we can H (or the pair (S, Ξ)) a *matroid*.

Examples: i) Matric matroid: A matrix $A = (a_1, \ldots, a_n)_{n\times m}$, $S = \{a_1, \ldots, a_n\}$,

$$X \in \Xi \Longleftrightarrow X = \{a_{i_1}, \ldots, a_{i_k}\} \text{ is independent.}$$

A special case of i) with $A =$ the vertex-edge incidence matrix.

ii) Graphic matroid: $G = (V, E)$, $S = E$,

$$X \in \Xi \Longleftrightarrow X \subseteq E,\ X \text{ has no cycle.}$$

ii) is a special case of i) with $A =$ the vertex-edge incidence matrix.

A Base := a maximal independence set; that is

$$B \in \Xi, \text{ no } X \supset B(X \neq B), \; X \in \Xi.$$

Proposition. If X and Y are two bases of a matroid $H = (S, \Xi)$, then $|X| = |Y|$.

Matroid Problem: $W : S \to \Re_+$. $\forall e \in S, W(e) \geq 0$ defined, $X \subseteq S$,

$$W(X) := \sum_{e \in X} W(e).$$

Then the matroid problem is

$$\begin{array}{ll} \max & W(X) \\ \text{s.t.} & X \in \Xi. \end{array}$$

Rank Function: $X \subseteq S$, $r(X) = \max\{|Y| : \; Y \in \Xi, Y \subseteq X\}$.

Circuit: $C \subseteq S$, $C \notin \Xi$. But any $X \subset C$ $(X \neq C)$, $\Longrightarrow X \in \Xi$.

Span: $\text{sp}(X) = Y : Y \supseteq X, \; r(Y) = r(X), \; \text{no } Z \supset Y(Z \neq Y), \, r(Z) = r(X)$.

Closed Set: $X \subseteq S$, $\forall e \in S \backslash X$, $r(X + e) = r(X) + 1$.

$$\Longrightarrow Y = \text{sp}(X) \Longrightarrow Y \text{ is closed.}$$

Theorem. $S_\rho(A) = \{e : \; r(A + e) = r(A)\}$ is unique.

7.2 More examples of matroids

1) Matching Matroid (*Edmonds and Fulkerson*):

$$G = (V, E), \; S \subseteq V$$

and

$$\Xi = \{X \subseteq S : \text{ there exists a matching } M \text{ covering } X\}.$$

Theorem. $H = (S, \Xi)$ is a matroid, called a *matching matroid*.
Proof. Clearly $\emptyset \in \Xi$ and $X \subseteq Y \in \Xi \Longrightarrow X \in \Xi$.

Now suppose that X and Y are two sets in Ξ containing r and $r+1$ vertices, respectively. Let M_X and M_Y be matchings covering X and Y, respectively. Assume that for all $v \in Y \backslash X$, v is not covered by M_X, else M_X covers $X + v$, for some $v \in Y \backslash X$, and the conclusion follows immediately. Consider the symmetric difference of the matchings M_X

and M_Y, which is composed of alternating cycles and alternating paths. At least one of the alternating paths must extend between a vertex not in X and a vertex $v \in Y \backslash X$. The symmetric difference of this alternating path and M_X yields a matching which covers $X + v$. Hence H is a matroid. $\square$

2) Transversal Matroid (*Edmonds and Fulkerson*):

$$S = \{e_1, \ldots, e_n\},\ T = \{t_1, \ldots, t_m\},\ t_j \subseteq S.$$

$X \subseteq S$ is called a partial transversal if $\forall\ e_i \in X$, there exists a unique t_j such that $e_i \in t_j$. Let

$$\Xi = \{X \subseteq S:\ X \text{ is a partial transversal}\}.$$

Let $G = (S, T, E)$ be bipartite graph where edge $(i, j) \in E$ if and only if $e_i \in t_j$. Then $H = (S, \Xi)$ is a matching matroid, which is called a *transversal matroid*.

3) Partition Matroid (*Edmonds and Fulkerson*):

$$S = \{e_1, \ldots, e_n\},\ m_1, \ldots, m_k \geq 0,\ \text{integer}.$$

Suppose that $\{S_1, \ldots, S_k\}$ is a partition of S (disjoint and exhaustive division into subsets). Let

$$\Xi = \{X \subseteq S:\ |X \cap S_i| \leq m_i,\ i = 1, \ldots, k\}.$$

Then $H = (S, \Xi)$ is a special case of the transversal matroid, which is called a *partition matroid*.

4) Tail Matroid $-$ Tail Partition Matroid:

$$G = (N, A),\ \text{a diagraph},\ S = A.$$

Let

$$\Xi = \{X \subseteq S:\ \text{for any } i \in N,\ X \text{ has at most one tail from } i\}$$

and

$$S_i = A_i = \{(i, j) \in A:\ j \in N\}.$$

Let $m_i = 1$ for $i = 1, \ldots, n$. Then for any $X \in \Xi$,

$$|X \cap S_i| \leq m_i,\ i = 1, \ldots, n.$$

Hence, $H = (S, \Xi)$ is a partition matroid $-$ *tail partition matroid*.

5) Head Matroid $-$ Head Partition Matroid:

$$G = (N, A), \text{ a diagraph, } S = A.$$

Let

$$\Xi = \{X \subseteq S: \text{ for any } i \in N, \ X \text{ has at most a head to } i\}$$

and

$$S_i = A_i = \{(j, i) \in A: \ j \in N\}.$$

Let $m_i = 1$ for $i = 1, \ldots, n$. Then for any $X \in \Xi$,

$$|X \cap S_i| \leq m_i, \ i = 1, \ldots, n.$$

Hence, $H = (S, \Xi)$ is a partition matroid $-$ *head partition matroid.*

7.3 Matroid Greedy Algorithms

Suppose that $H = (S, \Xi)$ is an independent system and $W: S \to \Re_+$ is a weight function with $W(e) \geq 0 \ \forall e \in S$. For $X \subseteq S$, let

$$W(X) := \sum_{e \in X} W(e).$$

Then the matroid problem is

$$\begin{array}{ll} \max & W(X) \\ \text{s.t.} & X \in \Xi. \end{array}$$

Greedy Algorithm $-$ Matroid Algorithm:

Suppose $W(e_1) \geq W(e_2) \geq \ldots \geq W(e_n)$.

Step 0. Let $X = \emptyset$.

Step k. If $X + e_k \in \Xi$, let $X := X + e_k$, where $k = 1, \ldots, n$.

Theorem. (Rado, Edmonds) The above algorithm works if and only if H is a matroid.

Applications:

1) The Maximal Spanning Tree Problem.

Suppose that there is a television network wishes to lease video links so that its stations in various places can be formed into a connected network. Each link (i, j) has a different rental cost c_{ij}. The question is how the network can be constructed to have the minimum cost? Obviously, what is wanted is a minimum cost spanning tree of video links. Replacing

c_{ij} by $M - c_{ij}$, where M is a larger number, we can see that it then turns into a *maximum spanning tree* (MST). Kruskal has already proposed the following solution: *Choose the edges one at a time in order of their weights, largest first, rejecting an arc only if it forms a cycle with edges already chosen.*

2) A Sequencing Problem.

Suppose that there are a number of jobs which are to be processed by a single machine. All jobs require the same processing time. Each job j has assigned to it a deadline d_j, and a penalty p_j, which must be paid if the job is not completed by its deadline. What ordering of the jobs minimizes the total penalty costs? It can be easily seen that there exists an optimal sequence in which all jobs completed on time appear at the beginning of the sequence in order of deadlines, earliest deadline first. The late jobs follow, in arbitrary order. Thus, the problem is to choose an optimal set of jobs which can be completed on time.

3) A Semimatching Problem.

Let W be an $m \times n$ nonnegative matrix. Suppose we wish to choose a maximum weight subset of elements, subject to the constraint that no two elements are from the same row of the matrix.

Matroid intersection problems, 3-matroid intersection problems and matroid matching problems.