Matroid theory (Cormen et. al., Chapter 17)

A matroid is an ordered pair M = (S, ℓ) satisfying the following conditions.

1. S is a finite nonempty set.
2. ℓ is a nonempty family of subsets of S, called the *independent subsets* of S, such that if B $\in \ell$ and A $\subseteq$ B, then A$\in\ell$. We say that ℓ is *hereditary* if it satisfies this property. Note that the empty set Ø is necessarily a member of ℓ.
3. If A $\in \ell$, B $\in \ell$, and |A| < |B|, then there is some element x $\in$ B - A such that A $\cup$ {x} $\in \ell$. We say that M satisfies the *exchange* property.

Some terminologies: A$\in\ell$ $\Rightarrow$ called independent; I$\notin \ell\Rightarrow$ dependent. A *circuit* is an inclusionwise minimally dependent set of S. A *basis* is any maximal independent set.
I is a spanning set if I $\supseteq$ B for some basis B.

Note: all bases of a matroid M must have the same cardinality (because of the hereditary property).

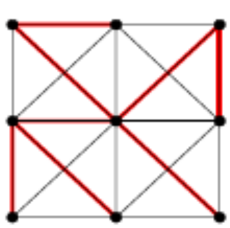
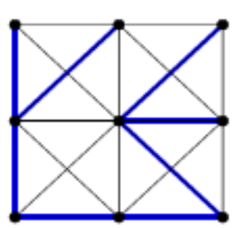
Example 1: (Uniform matroids) Given by |S| = n, ℓ = {I $\subseteq$ S : |I| $\leq$ k}. The circuits are all sets of cardinality k + 1, and the bases are all sets of cardinality k.

A graphic matroid $\mathbf{M_G}$ = ($\mathbf{S_G}$, $\boldsymbol{\ell_G}$) is defined in terms of a given (finite) undirected graph **G** = (**V**, **E**):

1. The set $\mathbf{S_G}$ is defined to be **E**, the set of edges of **G**.
2. If **A** is a subset of **E**, then **A** $\in \ell_G$ if and only if **A** is acyclic. That is, a set of edges **A** is *independent* if and only if the subgraph $\mathbf{G_A}$ = (**V**, **A**) forms a forest.

Theorem 1. If G = (V, E) is an undirected graph, then M_G = (S_G, ℓ_G) is a matroid.
Proof: $\mathbf{S_G}$ = **E** is a finite set. ℓ_G is hereditary, since a subset of a forest is a forest. Suppose that $\mathbf{G_A}$ = (**V, A**) and $\mathbf{G_B}$ = (**V, B**) are forests of **G** and that |**B**| > |**A**|. That is, **A** and **B** are acyclic sets of edges, and **B** contains more edges than **A** does. Forest $\mathbf{G_B}$ must contain an edge (**u, v**) such that vertices **u** and **v** are in different trees in forest $\mathbf{G_A}$. Then (**u, v**) can be added to forest $\mathbf{G_A}$ without creating a cycle, and thus $\mathbf{G_A}$ $\cup$ (**u, v**) $\in \ell_G$. Therefore, $\mathbf{M_G}$ satisfies the exchange property, completing the proof that $\mathbf{M_G}$ is a matroid.

Consider a graphic matroid $\mathbf{M_G}$ for a connected, undirected graph **G**. Every *basis* (maximal independent set of edges) must be a tree with exactly |**V**| - 1 edges that connects all the vertices of **G**, which is a *spanning tree* of G.

B   B'

Many problems for which a greedy approach provides optimal solutions can be formulated in terms of finding a maximum-weight independent (i.e. the *optimal*) subset in a weighted matroid. That is, we are given a weighted matroid $M = (S,\boldsymbol{\ell})$, and we wish to find an independent set $A \in \boldsymbol{\ell}$ such that $w(A)$ is maximized. Because the weight $w(x)$ of any element $x \in S$ is positive, an optimal subset is always a maximal independent subset.

<u>Example: Minimum spanning tree problem.</u>
We are given a connected undirected graph $G = (V, E)$ and a length function w such that $w(e)$ is the (positive) length of edge e. We are asked to find a subset of the edges that connects all of the vertices together and has minimum total length. To view this as a problem of finding an optimal subset of a matroid, consider the weighted matroid $\mathbf{M_G}$ with weight function $\mathbf{w'}$, where $\mathbf{w'(e) = w_0 - w(e)}$ and $\mathbf{w_0}$ is larger than the maximum length of any edge. In this weighted matroid, all weights are positive and an optimal subset is a spanning tree of minimum total length in the original graph. More specifically, each maximal independent subset **A** corresponds to a spanning tree and since $w'(A) = (|V| - 1)w_0 - w(A)$ for any maximal independent subset A, an independent subset that maximizes the quantity $w'(A)$ must minimize $w(A)$.

Below is a greedy algorithm that finds an optimal subset A in an arbitrary matroid. The algorithm takes as input a weighted matroid $M = (S,\boldsymbol{\ell})$ with an associated positive weight function w, and it returns an optimal subset A. In our pseudocode, we denote the components of M by S[M] and $\boldsymbol{\ell}$[M], and the weight function by w.

```
GREEDY(M, w)
1 A ← Ø
2 sort S [M ] into monotonically decreasing order by weight w
3 for  each x ∈ S[M], taken in monotonically decreasing order by weight w (x)
4          if A ∪ {x } ∈ ℓ [M]
5                    A ← A  ∪ {x}
6 return A
```

The entire algorithm runs in time $\mathbf{O(n \lg n + nf(n))}$, if each check in line 4 takes time $\mathbf{O(f(n))}$.

Since the empty set is independent by the definition of a matroid, and since x is added to A only if $A \cup \{x\}$ is independent, the subset A is always independent, by induction. Therefore, GREEDY always returns an independent subset A. The two lemma below shows that GREEDY gives an optimal independent subset.

Lemma. Suppose that $M = (S,\boldsymbol{\ell})$ is a weighted matroid with weight function w and that S is sorted into monotonically decreasing order by weight. Let x be the first element of S such that $\{x\}$ is independent, if any such x exists. If x exists, then there exists an optimal subset A of S that contains x.
Proof: If no such x exists, then the only independent subset is the empty set and we're done. Otherwise, let B be any nonempty optimal subset. Assume that $x \notin B$; otherwise, we let $A = B$ and we're done. No element of B has weight greater than $w(x)$. To see this, observe that $y \in B$ implies that $\{y\}$ is independent (because of the hereditary property),

and our choice of x therefore ensures that $w(x) \geq w(y)$. Construct the set A as follows. Begin with $A = \{x\}$. By the choice of x, A is independent. Using the exchange property, repeatedly find a new element of B that can be added to A until $|A| = |B|$ while preserving the independence of A. Then, $A = B - \{y\} \cup \{x\}$ for some $y \in B$. As $\mathbf{w(A) = w(B) - w(y) + w(x) \geq w(B)}$, and because B is optimal, A must also be optimal.

Given a matrod $M = (S,\ell)$, $x \in S$ is called an *extension* of an independent subset A, if $A \cup \{x\}$ is also an independent set.

Lemma. Let $M = (S,\ell)$ be any matroid. If x is an element of S that is an extension of some independent subset A of S, then x is also an extension of Ø.
Proof: notice the hereditary property of A.

This implies that if x x is not an extension of Ø, then x is not an extension of any independent subset A of S. Therefore, GREEDY cannot make an error by passing over any initial elements in S that are not an extension of Ø, since they can never be used.

Let x be the first element of S chosen by GREEDY for the weighted matroid $M = (S,\ell)$. The remaining problem of finding a maximum-weight independent subset containing x reduces to finding a maximum-weight independent subset of the weighted matroid $M' = (S',\ell)$, where $S' = \{y \in S: \{x, y\} \in \ell\}$, $\ell' = \{B \subseteq S-\{x\}: B \cup \{x\} \in \ell\}$, and the weight function for M' is the weight function for M, restricted to S'.

Therefore, GREEDY returns an optimal subset.

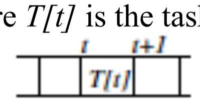
<u>Formulating a task-scheduling problem as a matroid</u>

<u>Instance</u>: A set $S = \{a_1, a_2, ..., a_n\}$ of n unit-time tasks (each requiring exactly one unit of time) to be scheduled where task i has deadline d_i, $1 \leq d_i \leq n$ (task a_i is supposed to finish by time d_i) and penalty w_i (applied if the task finish after d_i).

<u>Problem</u>: Determine a schedule for the tasks such that the sum of the penalties of the tasks not finished by their deadlines (not started by time d_i-1) is minimized

A schedule is specified by an array *T[0:n-1]* where *T[t]* is the task scheduled at time *t*.

The task started at time *t* is complete at time *t+1*.



A set A of tasks is *independent* if there exists a schedule for these tasks such that no tasks are late. Then the scheduling problem is equivalent to find an independent subset of tasks A to maximize the total weight of tasks in A (or to minimize the total weight of tasks in S-A).

Let $N_t(A)$ be the number of tasks in A whose deadline is t or earlier.

A set of tasks A is independent if and only if for all t, $N_t(A) \leq t$.

Theorem: Let S be the set of unit-time tasks with deadlines, and $\boldsymbol{\ell}$ be the set of all independent sets of tasks, then the corresponding system M = (S,$\boldsymbol{\ell}$) is a matroid.
Proof: Every subset of an independent set of tasks is certainly independent. To prove the exchange property, suppose that B and A are independent sets of tasks and that $|B| > |A|$. Let k be the largest t such that $N_t(B) \leq N_t(A)$. (Such a value of t exists, since $N_0(A) = N_0(B) = 0$.) Since $N_n(B) = |B|$ and $N_n(A)=|A|$, but $|B|>|A|$, we must have that $k < n$ and that $N_j(B)>N_j(A)$ for all j, $k+1 \leq j \leq n$. Therefore, there is a task with deadline k + 1 in B but not in A. Let a_i be the task in B-A with deadline k +1. Let $A' = A \cup \{a_i\}$. A' is independent, because for all $t \leq k$, $N_j(A') = N_j(A) \leq t$; for all $t \geq k+1$, $N_j(A') = N_j(A)+1 \leq N_j(B) \leq t$.

Therefore, an optimal independent subset can be obtained from

```
GREEDY_SCHEDULE(M, w)
A ← Ø
sort tasks by non-increasing weight w
for each task x, taken in non-increasing order by weight w(x)
        A' ← A∪{x}
        if A' satisfies N_t(A') ≤ t
                A ← A;
return A
```

```
GREEDY SCHEDULE(M, w)
initialize T by 0 // no task is scheduled
sort tasks by non-increasing weight w
for each task x, taken in non-increasing order by weight w(x) // x>0
        d ← d[x] – 1
        while (d ≥ 0 and T[d]≠0)        // d[x]-1 is already scheduled
                d ← d-1                 // try previous time point
        if d ≥ 0
                T[d] ← x                // schedule x at time d
Output T                                // full schedule of the optimal independent set
```

Computational complexity: $O(n^2)$