We first review our definitions:

A **subset system** is a set $E$ together with a *set of subsets* of $E$, called $I$, such that $I$ is **closed under inclusion**. This means that if $X \subseteq Y$ and $Y \in I$, then $X \in I$.

The **optimization problem** for a subset system $(E, I)$ has as input a positive weight for each element of $E$. Its output is a set $X \in I$ such that $X$ has at least as much total weight as any other set in $I$.

A subset system is a **matroid** if it satisfies the **exchange property**: If $i$ and $i'$ are sets in $I$ and $i$ has fewer elements than $i'$, then there exists an element $e \in i' \setminus i$ such that $i \cup \{e\} \in I$.

## The Generic Greedy Algorithm

Given any *finite* subset system $(E, I)$, we find a set in $I$ as follows:

- Set $X$ to $\emptyset$.
- Sort the elements of $E$ by weight, heaviest first.
- For each element of $E$ in this order, add it to $X$ iff the result is in $I$.
- Return $X$.

Today we prove:

**Theorem:** For any subset system $(E, I)$, the greedy algorithm solves the optimization problem for $(E, I)$ if and only if $(E, I)$ is a matroid.

**Theorem:** For any subset system $(E, I)$, the greedy algorithm solves the optimization problem for $(E, I)$ if and only if $(E, I)$ is a matroid.

**Proof:** We will show first that if $(E, I)$ is a matroid, then the greedy algorithm is correct.

Assume that $(E, I)$ satisfies the exchange property. Pick an arbitrary weight function, let $X$ be the set chosen by the greedy algorithm, and let $Y$ be any other maximal set. We show that $X$ has weight at least that of $Y$.

First note that $X$ and $Y$ must have the same size, which we will call $n$. If one had fewer elements, we could add an element of the other and stay in $I$. But $Y$ is assumed to be maximal in $I$, and if $X$ were not maximal the greedy algorithm would add an element to it rather than stopping with it.

(We are proving that if $(E, I)$ is a matroid, greedy set $X$ has weight at least that of any arbitrary maximal set $Y$.)

$$\begin{aligned} X &= \{x_1, \ldots, x_n\} \\ Y &= \{y_1, \ldots, y_n\} \end{aligned}$$

Here the elements are listed in descending order of weight. If $Y$ has total weight greater than that of $X$, then for some $k$, $w(x_k) < w(y_k)$. Choose the smallest such $k$.

Now let $\alpha$ be the first $k-1$ elements of $X$ and let $\beta$ be the first $k$ elements of $Y$. By the Exchange Property, we can make a set $Z$ in $I$ by adding one of the elements of $\beta$ to $\alpha$. Since each element of $\beta$ has weight greater than that of $x_k$, $Z$ has greater weight than $\{x_1, \ldots, x_k\}$.

This means that at some point, the greedy algorithm chose an $x_j$ when a higher-weight element of $Z$ was available. This contradicts the definition of the greedy algorithm.

Now we must prove that if $(E, I)$ fails to satisfy the Exchange Property, then there is some weight function on which the greedy algorithm fails.

Suppose there are two sets $i$ and $i'$ in $I$, with $|i| < |i'|$, such that no element of $i' \setminus i$ can be added to $i$ while keeping the result in $I$. Let $m$ be $|i|$. Our weight function is:

- Elements in $i$ have weight $m + 2$,
- Elements in $i' \setminus i$ have weight $m + 1$,
- Other elements have weight $0$.

**Greedy Algorithm:** Tries elements of weight $m+2$ first, gets all $m$ of them, then is stuck because no element of weight $m + 1$ fits, total score $m(m + 2)$.

**Optimal Algorithm:** Does at least as well as the set $i'$, which has total weight at least $m^2 + 2m + 1$ because each of its elements has weight at least $m + 1$.

This concludes both halves of the proof.

A subset system is a **matroid** if it satisfies the **exchange property**: If $i$ and $i'$ are sets in $I$ and $i$ has fewer elements than $i'$, then there exists an element $e \in i' \setminus i$ such that $i \cup \{e\} \in I$.

A subset system $(E, I)$ satisfies the **Cardinality Property** if for any set $A \subseteq E$, all maximal independent sets in $A$ have the same number of elements. ($X$ is a maximal independent set in $A$ if $X \in I$ and there is no set $Y \in I$ with $X \subsetneq Y \subseteq A$.)

**The Cardinality Theorem:** A subset system is a matroid iff it satisfies the Cardinality Property.

**The Cardinality Theorem:** A subset system is a matroid iff it satisfies the Cardinality Property.

**Proof:** We showed earlier that in a matroid, all sets that are *maximal in* $E$ must have the same cardinality, but now we must show a bit more. Let $A$ be a set and let $X$ and $Y$ be two sets in $I$ that are maximal in $A$. We must show that $X$ and $Y$ have the same size.

Suppose $X$ is smaller than $Y$. Then by the Exchange Property, we can add some element of $Y$ to $X$ and keep the result $Z$ in $I$. But since $X$ and $Y$ are both subsets of $A$, the set $Z$ is also a subset of $A$ and thus $X$ is not maximal in $A$.

**The Cardinality Theorem:** A subset system is a matroid iff it satisfies the Cardinality Property.

For the other half of the proof, we will show that if $(E, I)$ is *not* a matroid, then it fails to satisfy the Cardinality Property. Let $X$ and $Y$ be two sets in $I$ such that $|X| < |Y|$ but no element of $Y \setminus X$ can be added to $X$ to get a result in $I$.

We let $A$ be $X \cup Y$. Now $X$ is a maximal set in $A$, since we cannot add any of the other elements of $A$ to it.

The set $Y$ may *not* be maximal in $A$, but if it isn't there is some subset of $A$ that contains it and is maximal in $A$. Since this set is at least as big as $Y$, it is strictly bigger than $X$ and we have a violation of the Cardinality Property.

If we now go back to the Maximum Weight Forest problem, we can see fairly easily that the subset system $(E, I)$, where $E$ consists of the edges of the graph $G$ and $I$ consists of the acyclic sets of edges, satisfies the Cardinality Property.

If $G$ is connected, the maximal sets in $I$ must be spanning trees. If not, they are spanning forests – forests that consist of a spanning tree for each connected component of $G$. (This is because if there were two nodes connected by a path in $G$ but not in the forest, there would be an edge we could add to the forest without creating a cycle.

But we proved earlier that any forest on $n$ nodes with $c$ connected components, it has exactly $n - c$ edges. If $A$ is any set of edges, the subgraph of $G$ with edge set $A$ has *some* number of connnected components, $c$. Any maximal acyclic set of edges must have exactly $n - c$ edges. So all the maximal subsets of $A$ have cardinality $n - c$. and the Cardinality Property holds.

So the acyclic-edge-set subset system is a matroid, and thus our general results about matroids *prove* that the Kruskal algorithm always produces a minimum spanning tree. (As the greedy algorithm for MWF on the related weight function would give us a maximum weight forest whose edges would be an MST of the original graph.)

Next time we'll consider the **maximum matching** problem, where we saw that the greedy algorithm does not always work. The corresponding subset system is not a matroid, but we will see that it is the *intersection* of two matroids. We'll present an algorithm for the matching problem on bipartite graphs, and show how it can be adapted to any system that is the intersection of two matroids.