Matroids and greedy algorithms

Recall minimum spanning trees and Kruskal's algorithm.

Let $G=(V,E,w)$ when $w: E \to \mathbb{R}_0$ $(w(e) \geq 0 \ \forall e \in E)$ be given and let $W = \max\{w(e) \mid e \in E\}$

If we set $w'(e) = W - w(e)$ then for any spanning tree T of G we have $(n = |V|)$

$$w'(T) = (n-1)W - w(T)$$

Hence T is a minimum spanning tree wrt w

$$\Updownarrow$$

T is a maximum weigh spanning tree wrt w'

From now on we look only at the maximization version (maximum weigh spanning trees)

Greedy algorithm for MST (Kruskal)

$F \leftarrow \emptyset$

while $E \neq \emptyset$

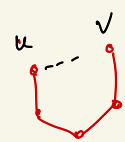
  Let $uv \in E$ have maximum weight

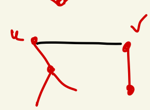
  $E \leftarrow E - uv$

  If u and v are not connected via F

    $F \leftarrow F + uv$

end





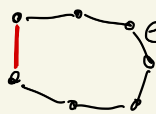
Theorem The greedy algorithm finds a maximum weight spanning tree in $G=(V,E,w)$

Proof: Let T_g be the greedy spanning tree and suppose $w(T_g) < w(T_{opt})$ where T_{opt} is a max weight spT of G.

Let $e_1, e_2, \dots e_{n-1}$ be the edges that we add when T_g is constructed (in that order).

Choose T^* max weight spT and j s.t. $\{e_1, e_2, \dots e_j\} \subseteq E(T^*)$ and no max weight (*) spT contains all the edges $\{e_1, \dots e_{j+1}\}$

$T^* + e_{j+1}$ contains a cycle C



At least one edge of C is not in T_g as this is a tree.

Let e be such an edge.

Then $e \in E - \{e_1, e_2, \dots e_j, e_{j+1}\}$ so $w(e_{j+1}) \ge w(e)$ since the greedy algorithm chose e_{j+1} in step $j+1$

Now $w(T^*) \ge w(T^* - e + e_{j+1}) \ge w(T^*)$ so $T' = T^* - e + e_{j+1}$ is a max weight spT which violates (*) ↯

□.

Definition Let S be a finite set and $\mathcal{I} \subseteq 2^S$ a collection of subsets of S. Then the pair $M = (S, \mathcal{I})$ is a matroid if the following holds

(1) $\emptyset \in \mathcal{I}$

(2) If $X \in \mathcal{I}$ and $Y \subseteq X$ then $Y \in \mathcal{I}$

(3) if $X, Y \in \mathcal{I}$ and $|Y| = |X| + 1$ then $\exists y \in Y - X$ s.t $X + y \in \mathcal{I}$

(2) says that $\mathcal{I}$ is hereditary

A set Z s.t $Z \notin \mathcal{I}$ is dependent

A set X s.t $X \in \mathcal{I}$ is independent

Examples of Matroids

(A) $G = (V, E)$ a graph $S = E$

$\mathcal{I} = \{E' \subseteq E \mid E'$ induces a forest in $G\}$

This is called the cycle matroid of G

(B) S finite subset of a vector space V

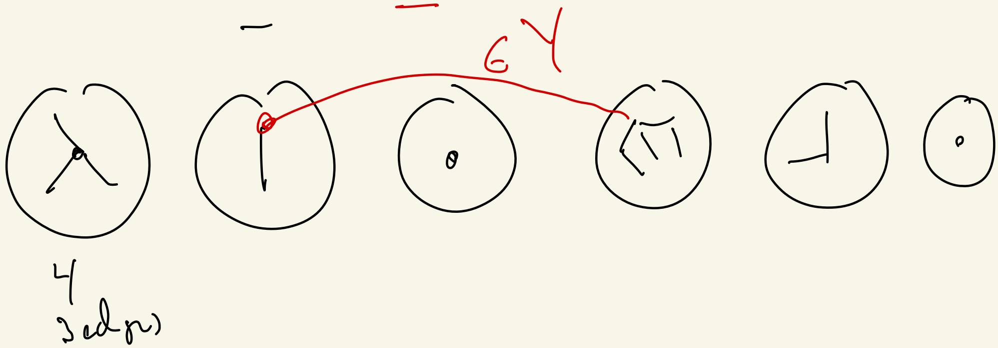
say that $\{x_1, x_2 \ldots x_k\} \in \mathcal{I}$ iff the vectors $x_1, x_2 \ldots x_k$ are linearly independent

Why does (3) hold when

$\mathcal{I} = \{E' \subseteq E \mid E' \text{ acyclic}\}$?

Suppose $X \subseteq E$ and $Y \subseteq E$

have $|X| = |Y| - 1$ so $|Y| = |X| + 1$



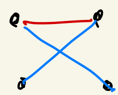
Property (3) of Matroids easily implies the following property which is called the exchange property

(3') if $X, Y \in \mathcal{F}$ and $|Y| = |X| + k$ for some $k \geq 1$ then there are elements $y_1, y_2 \dots y_k \in Y - X$ such that

$$X + y_1 \in \mathcal{F}, \quad X + y_1 + y_2 \in \mathcal{F}, \quad \dots, \quad X + y_1 + \dots + y_k \in \mathcal{F}$$

A base of a matroid $M = (S, \mathcal{F})$ is a maximal independent set B, that is $B \in \mathcal{F}$ but $B + w \notin \mathcal{F}$ $\forall w \in S - B$

Matchings do not form a matroid



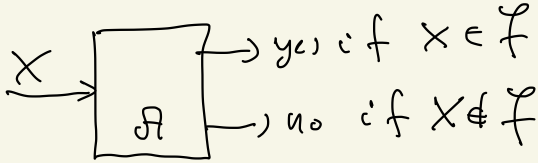
(3) does not hold

Corollary

All the bases of a matroid have the same size

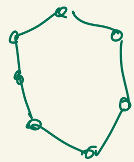
How do we check whether $X \in \mathcal{F}$?

We assume $M = (S, \mathcal{F})$ is given as S together with an oracle $\mathcal{A}$



Circuits of matroids

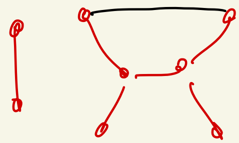
Cycle:



A circuit of a matroid $M=(S,\mathcal{F})$ is a minimally dependent subset of S:

Z is a circuit of $M \iff Z\notin\mathcal{F}$ and $Z-a\in\mathcal{F}\ \forall a\in Z$

in graphs



Proposition: Let $M=(S,\mathcal{F})$ be a matroid. If $X\in\mathcal{F}$ but $X+y\notin\mathcal{F}$ then there is a unique circuit C contained in $X+y$

proof:

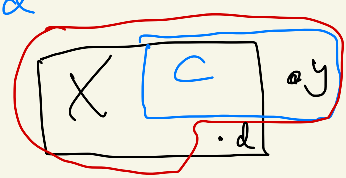
$X+y$ is dependent so it contains at least one circuit

let $C=\{a\in X+y \mid X+y-a\in\mathcal{F}\}$

claim: C is a circuit of M

proof: As $C-a\subseteq X+y-a\in\mathcal{F}$ $\forall a\in C$ implying that $C-a\in\mathcal{F}$ we just need to show that $C\notin\mathcal{F}$.

Suppose $C\in\mathcal{F}$. Then by (3') we can add elements from $X-C$ to C until we have a set $Y\in\mathcal{F}$ with $Y\in\mathcal{F}$ and $|Y|=|X|$. But then $Y=X+y-d$ for some d implying that $d\in C$ by def of C

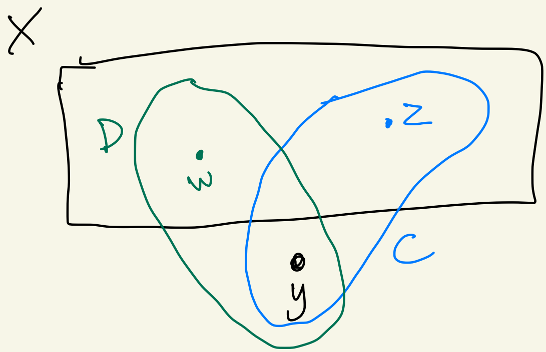


contradicting that $C\subseteq Y\in\mathcal{F}$

We proved that C is a circuit so it only remains to prove that C is the unique circuit in $X+y$.

Suppose $D \subseteq X+y$ is another circuit

Then the picture is



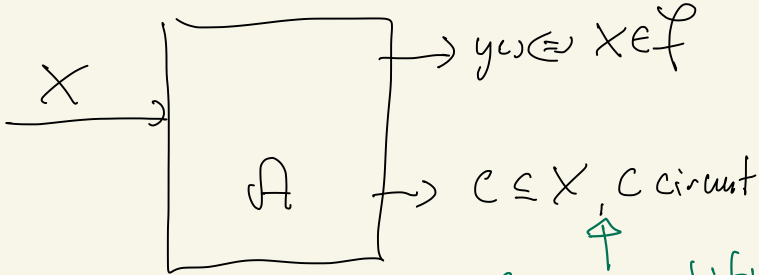
By minimality of circuits $C \nsubseteq D$

So $\exists z \in C - D$ (and $\exists w \in D - C$)

By definition of C the set $X+y-z \in \mathcal{I}$ but $D \subseteq X+y-z$ contradicts (2) in the definition of matroids $\square$.

The greedy algorithm for matroids

Below we assume that the matroid $M=(S,\mathcal{F})$ is given in form of an oracle $\mathcal{A}$ such that

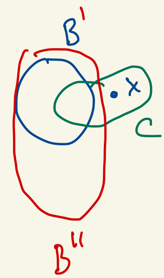


C is a certificate that $X \notin \mathcal{F}$

Greedy algorithm:

input: a matroid $M=(S,\mathcal{F})$ and $w: S \to \mathbb{R}_{+}\cup\{0\}$
output: a maximum weight basis of M

$B \leftarrow \emptyset$, $S' \leftarrow S$
while $S' \neq \emptyset$ do
  let $x \in S'$ have $w(x) \geq w(y)$ $\forall y \in S'$
  $S' \leftarrow S' - x$
  if $B+x \in \mathcal{F}$ then $B \leftarrow B+x$
end
output B



Theorem The greedy algorithm outputs a maximum weight basis of M.

Proof Let B_g denote the output and let B^* be an optimum basis $(w(B^*) \geq w(B) \ \forall$ basis $B)$

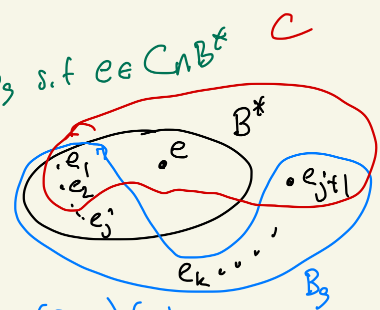
Let $e_1, e_2, \dots, e_k$ be the elements added to B by the algorithm in that order.

Let $j \in [k]$ be maximum s.t. $\{e_1, e_2, \dots, e_j\} \subseteq B^*$ for some optimum basis.

If $j = k$ we are done (B_g is optimal) so assume $j < k$

Then $B^* + e_{j+1}$ contains a circuit C.

$C \nsubseteq B_g$ as C is dependent so $\exists e \in B^* - B_g$ s.t. $e \in C \cap B^*$



By the proposition, C is the unique circuit in $B^* + e_{j+1}$ so $\hat{B} = B^* + e_{j+1} - e$ is a basis

B^* is optimal so $w(B^*) \geq w(\hat{B}) \Rightarrow w(e) \geq w(e_{j+1})$ $\boxed{\text{I}}$

Recall that $\{e_1, e_2, \dots, e_j\} \subseteq B_g \cap B^*$ and $A = \{e_1, e_2, \dots, e_j, e\} \subseteq B^*$ so A is independent and hence e is a possible choice for the greedy algorithm in step $j+1$.

This implies that $w(e_{j+1}) \geq w(e)$ so by $\boxed{\text{I}}$ we have $w(e) = w(e_{j+1})$ and $w(\hat{B}) = w(B^*)$, contradicting the choice of B^* as $\{e_1, e_2, \dots, e_{j+1}\} \subseteq \hat{B}$

$\square$.

Alternative definition of a matroid (Def 12.2 in PS)

Definition The subset system $M = (E, \mathcal{I})$ is a matroid

$\Updownarrow$ def

The greedy algorithm finds an optimum basis for every non-negative weight function $\omega : E \to \mathbb{R}_+ \cup \{0\}$

Theorem 12.5 in PS Let $M = (E, \mathcal{I})$ be a subset system

(1) M is a matroid

$\Updownarrow$

(2) $\forall I_p, I_{p+1} \in \mathcal{I} : |I_p| = p = |I_{p+1}| - 1$
$\exists e \in I_{p+1} - I_p : I_p + e \in \mathcal{I}$

$\Updownarrow$

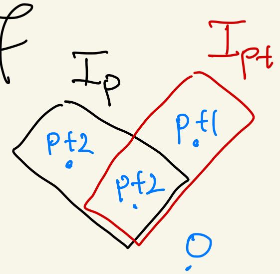
(3) If $A \subseteq E$ and I, I' are maximal independent subsets of A then $|I| = |I'|$

Proof

(1) $\Rightarrow$ (2):

suppose $I_p, I_{p+1} \in \mathcal{I}$; $|I_p| = p = |I_{p+1}| - 1$

but $\nexists\, e \in I_{p+1} - I_p$; $I_p + e \in \mathcal{I}$



let $w(e) = \begin{cases} p+2 & \text{if } e \in I_p \\ p+1 & \text{if } e \in I_{p+1} - I_p \\ 0 & \text{otherwise} \end{cases}$

Then $w(I_p) = p \cdot (p+2) < (p+1)^2 \leq w(I_{p+1})$

but G.A chooses I_p plus some 0 weight element $\rightarrow\leftarrow$ that G.A. works for w.

(2) $\Rightarrow$ (3) suppose I, I' are maximal independent subsets of A but $|I'| > |I|$

Then let $I'' \subseteq I'$ have $|I''| = |I| + 1$ and apply (2) to get that I is not a maximal subset of A }

(3) ⟹ (1): Suppose (3) holds but there is some non-negative w such that the G.A. does not find an optimal base.

Let $I = \{e_1, e_2, \dots e_i\}$ be the greedy solution where $w(e_1) \geq w(e_2) \geq \dots \geq w(e_i)$ and

$J = \{e_1', e_2' \dots e_j'\}$ be an optimal solution where $w(e_1') \geq w(e_2') \geq \dots \geq w(e_j')$.

Then $w(J) > w(I)$ by the assumption above

I is maximal by construction (via G.A.) and we can assume J is also maximal (by adding zero or more weight zero elements)

By (3) we have $i = |I| = |J|$

It suffices to prove the following claim

Claim $w(e_m) \geq w(e_m')$ for all $m = 1, 2, \dots, i$

proof of claim: $w(e_1) \geq w(e_1')$ as G.A chose e_1

suppose $\exists m \leq i$ s.t $w(e_j) \geq w(e_j')$ for $j = 1, 2, \dots, m-1$ but $w(e_m) < w(e_m')$

Let $A = \{e \in E \mid w(e) \geq w(e_m')\}$ then

$Z = \{e_1, e_2, \dots, e_{m-1}\} \subseteq A$ and Z is a maximal subset of A (as G.A chose $e_m \notin A$)

But $Z' = \{e_1', e_2' \dots, e_m'\} \subseteq A$ so it is contained in some maximal subset Z'' of A

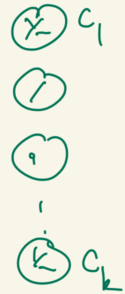
However now $|Z| < |Z''|$ $\rightarrow \leftarrow$ (3) $\square$.

Further definitions and remarks

Let $M = (E, \mathcal{I})$ be a matroid

The **rank** of $A \subseteq E$, denoted $r(A)$

is $r(A) = \max\{|X| \mid X \subseteq A \text{ and } X \in \mathcal{I}\}$



The rank of M is the size of a base of M (recall that they all have the same size)

$n-k$

$n - \#\text{comp}$

The set of bases of M is

$\mathcal{B}(M) = \{B \mid B \text{ is a base of } M\}$

Note that $\mathcal{I} = \{X \mid X \subseteq B \text{ for some } B \in \mathcal{B}(M)\}$

So $M = (E, \mathcal{I})$ is uniquely determined by its set of bases.