Submodular Functions, Optimization, and Applications to Machine Learning

— Spring Quarter, Lecture 9 —

http://www.ee.washington.edu/people/faculty/bilmes/classes/ee596b_spring_2016/

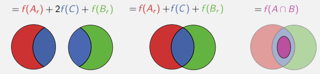
Prof. Jeff Bilmes

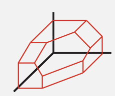
University of Washington, Seattle
Department of Electrical Engineering
http://melodi.ee.washington.edu/~bilmes

Apr 27, 2016

$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$

$= f(A_r) + 2f(C) + f(B_r)$ $\quad = f(A_r) + f(C) + f(B_r)$ $\quad = f(A \cap B)$





Cumulative Outstanding Reading

- Read chapters 2 and 3 from Fujishige's book.
- Read chapter 1 from Fujishige's book.

Announcements, Assignments, and Reminders

- Homework 3, available at our assignment dropbox (`https://canvas.uw.edu/courses/1039754/assignments`), due (electronically) Monday (5/2) at 11:55pm.
- Homework 2, available at our assignment dropbox (`https://canvas.uw.edu/courses/1039754/assignments`), due (electronically) Monday (4/18) at 11:55pm.
- Homework 1, available at our assignment dropbox (`https://canvas.uw.edu/courses/1039754/assignments`), due (electronically) Friday (4/8) at 11:55pm.
- Weekly Office Hours: Mondays, 3:30-4:30, or by skype or google hangout (set up meeting via our our discussion board (`https://canvas.uw.edu/courses/1039754/discussion_topics`)).

Class Road Map - IT-I

- L1(3/28): Motivation, Applications, & Basic Definitions
- L2(3/30): Machine Learning Apps (diversity, complexity, parameter, learning target, surrogate).
- L3(4/4): Info theory exs, more apps, definitions, graph/combinatorial examples, matrix rank example, visualization
- L4(4/6): Graph and Combinatorial Examples, matrix rank, Venn diagrams, examples of proofs of submodularity, some useful properties
- L5(4/11): Examples & Properties, Other Defs., Independence
- L6(4/13): Independence, Matroids, Matroid Examples, matroid rank is submodular
- L7(4/18): Matroid Rank, More on Partition Matroid, System of Distinct Reps, Transversals, Transversal Matroid,
- L8(4/20): Transversals, Matroid and representation, Dual Matroids,
- L9(4/25): Dual Matroids, Properties, Combinatorial Geometries, Matroid and Greedy, Polyhedra, Matroid Polytopes
- L10(4/27):
- L11(5/2):
- L12(5/4):
- L13(5/9):
- L14(5/11):
- L15(5/16):
- L16(5/18):
- L17(5/23):
- L18(5/25):
- L19(6/1):
- L20(6/6): Final Presentations maximization.

Finals Week: June 6th-10th, 2016.

Spanning Sets

- We have the following definitions:

Definition 9.2.4 (spanning set of a set)

Given a matroid $\mathcal{M} = (V, \mathcal{I})$, and a set $Y \subseteq V$, then any set $X \subseteq Y$ such that $r(X) = r(Y)$ is called a spanning set of Y.

Definition 9.2.5 (spanning set of a matroid)

Given a matroid $\mathcal{M} = (V, \mathcal{I})$, any set $A \subseteq V$ such that $r(A) = r(V)$ is called a spanning set of the matroid.

- A base of a matroid is a minimal spanning set (and it is independent) but supersets of a base are also spanning.
- V is always trivially spanning.
- Consider the terminology: "spanning tree in a graph", comes from spanning in a matroid sense.

Dual of a Matroid

- Given a matroid $M = (V, \mathcal{I})$, a dual matroid $M^* = (V, \mathcal{I}^*)$ can be defined on the same ground set V, but using a very different set of independent sets $\mathcal{I}^*$.
- We define the set of sets $\mathcal{I}^*$ for M^* as follows:

$$\mathcal{I}^* = \{A \subseteq V : V \setminus A \text{ is a spanning set of } M\} \quad (9.11)$$
$$= \{V \setminus S : S \subseteq V \text{ is a spanning set of } M\} \quad (9.12)$$

i.e., $\mathcal{I}^*$ are complements of spanning sets of M.
- That is, a set A is independent in the dual matroid M^* if removal of A from V does not decrease the rank in M:

$$\mathcal{I}^* = \{A \subseteq V : \text{rank}_M(V \setminus A) = \text{rank}_M(V)\} \quad (9.13)$$

- In other words, a set $A \subseteq V$ is independent in the dual M^* (i.e., $A \in \mathcal{I}^*$) if its complement is spanning in M (residual $V \setminus A$ must contain a base in M).
- Dual of the dual: Note, we have that $(M^*)^* = M$.

Dual of a Matroid: Bases

- The smallest spanning sets are bases. Hence, a base B of M (where $B = V \setminus B^*$ is as small as possible while still spanning) is the complement of a base B^* of M^* (where $B^* = V \setminus B$ is as large as possible while still being independent).
- In fact, we have that

Theorem 9.2.4 (Dual matroid bases)

Let $M = (V, \mathcal{I})$ be a matroid and $\mathcal{B}(M)$ be the set of bases of M. Then define

$$\mathcal{B}^*(M) = \{V \setminus B : B \in \mathcal{B}(M)\}. \tag{9.11}$$

Then $\mathcal{B}^*(M)$ is the set of basis of M^* (that is, $\mathcal{B}^*(M) = \mathcal{B}(M^*)$.

Example: cocycle matroid (sometimes "cut matroid")

- The dual of the cycle matroid is called the cocycle matroid. Recall, $\mathcal{I}^* = \{A \subseteq V : V \setminus A$ is a spanning set of $M\}$
- It consists of all sets of edges the complement of which contains a spanning tree — i.e., an independent set can't consist of edges that, if removed, would render the graph non-spanning.

The dual of a matroid is (indeed) a matroid

Theorem 9.3.1

Given matroid $M = (V, \mathcal{I})$, let $M^* = (V, \mathcal{I}^*)$ be as previously defined. Then M^* is a matroid.

Proof.

- Clearly $\emptyset \in \mathcal{I}^*$, so (I1') holds.
- Also, if $I \subseteq J \in \mathcal{I}^*$, then clearly also $I \in \mathcal{I}^*$ since if $V \setminus J$ is spanning in M, so must $V \setminus I$. Therefore, (I2') holds.
- Next, given $I, J \in \mathcal{I}^*$ with $|I| < |J|$, it must be the case that $\bar{I} = V \setminus I$ and $\bar{J} = V \setminus J$ are both spanning in M with $|\bar{I}| > |\bar{J}|$.

...

The dual of a matroid is (indeed) a matroid

Theorem 9.3.1

Given matroid $M = (V, \mathcal{I})$, let $M^* = (V, \mathcal{I}^*)$ be as previously defined. Then M^* is a matroid.

Proof.

- Consider $I, J \in \mathcal{I}^*$ with $|I| < |J|$. We need to show that there is some member $v \in J \setminus I$ such that $I + v$ is independent in M^*, which means that $V \setminus (I + v) = (V \setminus I) \setminus v = \bar{I} - v$ is still spanning in M. That is, removing v from $V \setminus I$ doesn't make $(V \setminus I) \setminus v$ not spanning in M.
- Since $V \setminus J$ is spanning in M, $V \setminus J$ contains some base (say $B_{\bar{J}} \subseteq V \setminus J$) of M. Also, $V \setminus I$ contains a base of M, say $B_{\bar{I}} \subseteq V \setminus I$.
- Since $B_{\bar{J}} \setminus I \subseteq V \setminus I$, and $B_{\bar{J}} \setminus I$ is independent in M, we can choose the base $B_{\bar{I}}$ of M s.t. $B_{\bar{J}} \setminus I \subseteq B_{\bar{I}} \subseteq V \setminus I$.
- Since $B_{\bar{J}}$ and J are disjoint, we have both: 1) $B_{\bar{J}} \setminus I$ and $J \setminus I$ are disjoint; and 2) $B_{\bar{J}} \cap I \subseteq I \setminus J$. Also note, $B_{\bar{I}}$ and I are disjoint.

...

The dual of a matroid is (indeed) a matroid

Theorem 9.3.1

Given matroid $M = (V, \mathcal{I})$, let $M^* = (V, \mathcal{I}^*)$ be as previously defined. Then M^* is a matroid.

Proof.

- Now $J \setminus I \not\subseteq B_{\bar{I}}$, since otherwise (i.e., assuming $J \setminus I \subseteq B_{\bar{I}}$):

$$
\begin{aligned}
|B_{\bar{J}}| &= |B_{\bar{J}} \cap I| + |B_{\bar{J}} \setminus I| && (9.1)\\
&\leq |I \setminus J| + |B_{\bar{J}} \setminus I| && (9.2)\\
&< |J \setminus I| + |B_{\bar{J}} \setminus I| \leq |B_{\bar{I}}| && (9.3)
\end{aligned}
$$

  which is a contradiction. *The last inequality on the right follows since $J \setminus I \subseteq B_{\bar{I}}$ (by assumption) and $B_{\bar{J}} \setminus I \subseteq B_{\bar{I}}$ implies that $(J \setminus I) \cup (B_{\bar{J}} \setminus I) \subseteq B_{\bar{I}}$, but since J and $B_{\bar{J}}$ are disjoint, we have that $|J \setminus I| + |B_{\bar{J}} \setminus I| \leq |B_{\bar{I}}|$.*
- Therefore, $J \setminus I \not\subseteq B_{\bar{I}}$, and there is a $v \in J \setminus I$ s.t. $v \notin B_{\bar{I}}$.
- So $B_{\bar{I}}$ is disjoint with $I \cup \{v\}$, means $B_{\bar{I}} \subseteq V \setminus (I \cup \{v\})$, or

Dual Matroid | Other Matroid Properties | Combinatorial Geometries | Matroid and Greedy | Polyhedra | Matroid Polytopes

Matroid Duals and Representability

Theorem 9.3.2

Let M be an $\mathbb{F}$-representable matroid (i.e., one that can be represented by a finite sized matrix over field $\mathbb{F}$). Then M^* is also $\mathbb{F}$-representable.

Hence, for matroids as general as matric matroids, duality does not extend the space of matroids that can be used.

Theorem 9.3.3

Let M be a graphic matroid (i.e., one that can be represented by a graph $G = (V, E)$). Then M^* is not necessarily also graphic.

Hence, for graphic matroids, duality can increase the space and power of matroids, and since they are based on a graph, they are relatively easy to use: 1) all cuts are dependent sets; 2) minimal cuts are cycles; 3) bases are one edge less than minimal cuts; and 4) independent sets are edges that are not cuts (minimal or otherwise).

Dual Matroid Rank

Theorem 9.3.4

The rank function r_{M^*} of the dual matroid M^* may be specified in terms of the rank r_M in matroid M as follows. For $X \subseteq V$:

$$r_{M^*}(X) = |X| + r_M(V \setminus X) - r_M(V) \tag{9.4}$$

- Note, we again immediately see that this is submodular by the properties of submodular functions we saw in lectures 1 and 2. *I.e., $|X|$ is modular, complement $f(V \setminus X)$ is submodular if f is submodular, $r_M(V)$ is a constant, and summing submodular functions and a constant preserves submodularity.*
- Non-negativity integral follows since $|X| + r_M(V \setminus X) \geq r_M(X) + r_M(V \setminus X) \geq r_M(V)$. *The right inequality follows since r_M is submodular.*
- Monotone non-decreasing follows since, as X increases by one, $|X|$ always increases by 1, while $r_M(V \setminus X)$ decreases by one or zero.
- Therefore, r_{M^*} is the rank function of a matroid. That it is the dual

Dual Matroid Rank

Theorem 9.3.4

The rank function r_{M^*} of the dual matroid M^* may be specified in terms of the rank r_M in matroid M as follows. For $X \subseteq V$:

$$r_{M^*}(X) = |X| + r_M(V \setminus X) - r_M(V) \tag{9.4}$$

Proof.

A set X is independent in (V, r_{M^*}) if and only if

$$r_{M^*}(X) = |X| + r_M(V \setminus X) - r_M(V) = |X| \tag{9.5}$$

or

$$r_M(V \setminus X) = r_M(V) \tag{9.6}$$

But a subset X is independent in M^* only if $V \setminus X$ is spanning in M (by the definition of the dual matroid). □

Matroid restriction/deletion

- Let $M = (V, \mathcal{I})$ be a matroid and let $Y \subseteq V$, then

$$\mathcal{I}_Y = \{Z : Z \subseteq Y, Z \in \mathcal{I}\} \tag{9.7}$$

  is such that $M_Y = (Y, \mathcal{I}_Y)$ is a matroid with rank $r(M_Y) = r(Y)$.
- This is called the restriction of M to Y, and is often written $M|Y$.
- If $Y = V \setminus X$, then we have that $M|Y$ has the form:

$$\mathcal{I}_Y = \{Z : Z \cap X = \emptyset, Z \in \mathcal{I}\} \tag{9.8}$$

  is considered a deletion of X from M, and is often written $M \setminus X$.
- Hence, $M|Y = M \setminus (V \setminus Y)$, and $M|(V \setminus X) = M \setminus X$.
- The rank function is of the same form. I.e., $r_Y : 2^Y \to \mathbb{Z}_+$, where $r_Y(Z) = r(Z)$ for $Z \subseteq Y$.

Matroid contraction M/Z

- Contraction by Z is dual to deletion, and is like a forced inclusion of a contained base B_Z of Z, but with a similar ground set removal by Z. Contracting Z is written M/Z. Updated ground set in M/Z is $V \setminus Z$.
- Let $Z \subseteq V$ and let B_Z be a base of Z. Then a subset $I \subseteq V \setminus Z$ is independent in M/Z iff $I \cup B_Z$ is independent in M.
- The rank function takes the form

$$r_{M/Z}(Y) = r(Y \cup Z) - r(Z) = r(Y|Z) \tag{9.9}$$

$$= r(Y \cup B_Z) - r(B_Z) \tag{9.10}$$

- So given $I \subseteq V \setminus Z$ and B_Z is a base of Z, $r_{M/Z}(I) = |I|$ is identical to $r(I \cup Z) = |I| + r(Z) = |I| + |B_Z|$ but $r(I \cup Z) = r(I \cup B_Z)$. This implies $r(I \cup B_Z) = |I| + |B_Z|$, or $I \cup B_Z$ is independent in M.
- A minor of a matroid is any matroid obtained via a series of deletions and contractions of some matroid.
- In fact, it is the case $M/Z = (M^* \setminus Z)^*$ (Exercise: show why).

Matroid Intersection

- Let $M_1 = (V, \mathcal{I}_1)$ and $M_2 = (V, \mathcal{I}_2)$ be two matroids. Consider their common independent sets $\mathcal{I}_1 \cap \mathcal{I}_2$.
- While $(V, \mathcal{I}_1 \cap \mathcal{I}_2)$ is typically not a matroid (Exercise: show graphical example.), we might be interested in finding the maximum size common independent set. That is, find $\max |X|$ such that both $X \in \mathcal{I}_1$ and $X \in \mathcal{I}_2$.

Theorem 9.4.1

Let M_1 and M_2 be given as above, with rank functions r_1 and r_2. Then the size of the maximum size set in $\mathcal{I}_1 \cap \mathcal{I}_2$ is given by

$$(r_1 * r_2)(V) \triangleq \min_{X \subseteq V} \Big(r_1(X) + r_2(V \setminus X) \Big) \tag{9.11}$$

This is an instance of the convolution of two submodular functions, f_1 and f_2 that, evaluated at $Y \subseteq V$, is written as:

$$(f_1 * f_2)(Y) = \min_{X \subseteq Y} \Big(f_1(X) + f_2(Y \setminus X) \Big) \tag{9.12}$$

Convolution and Hall's Theorem

- Recall Hall's theorem, that a transversal exists iff for all $X \subseteq V$, we have $|\Gamma(X)| \geq |X|$.
- $\Leftrightarrow \quad |\Gamma(X)| - |X| \geq 0, \forall X$
- $\Leftrightarrow \quad \min_X |\Gamma(X)| - |X| \geq 0$
- $\Leftrightarrow \quad \min_X |\Gamma(X)| + |V| - |X| \geq |V|$
- $\Leftrightarrow \quad \min_X \Big(|\Gamma(X)| + |V \setminus X| \Big) \geq |V|$
- $\Leftrightarrow \quad [\Gamma(\cdot) * |\cdot|](V) \geq |V|$
- So Hall's theorem can be expressed as convolution. Exercise: define $g(A) = [\Gamma(\cdot) * |\cdot|](A)$, prove that g is submodular.
- Note, in general, convolution of two submodular functions does not preserve submodularity (but in certain special cases it does).

Matroid Union

Definition 9.4.2

Let $M_1 = (V_1, \mathcal{I}_1)$, $M_2 = (V_2, \mathcal{I}_2)$, $\dots$, $M_k = (V_k, \mathcal{I}_k)$ be matroids. We define the union of matroids as $M_1 \vee M_2 \vee \dots \vee M_k = (V_1 \uplus V_2 \uplus \dots \uplus V_k, \mathcal{I}_1 \vee \mathcal{I}_2 \vee \dots \vee \mathcal{I}_k)$, where

$$I_1 \vee \mathcal{I}_2 \vee \dots \vee \mathcal{I}_k = \{I_1 \uplus I_2 \uplus \dots \uplus I_k | I_1 \in \mathcal{I}_1, \dots, I_k \in \mathcal{I}_k\} \tag{9.13}$$

Note $A \uplus B$ designates the disjoint union of A and B.

Theorem 9.4.3

Let $M_1 = (V_1, \mathcal{I}_1)$, $M_2 = (V_2, \mathcal{I}_2)$, $\dots$, $M_k = (V_k, \mathcal{I}_k)$ be matroids, with rank functions $r_1, \dots, r_k$. Then the union of these matroids is still a matroid, having rank function

$$r(Y) = \min_{X \subseteq Y} \Big(|Y \setminus X| + r_1(X \cap V_1) + \dots + r_k(X \cap V_k) \Big) \tag{9.14}$$

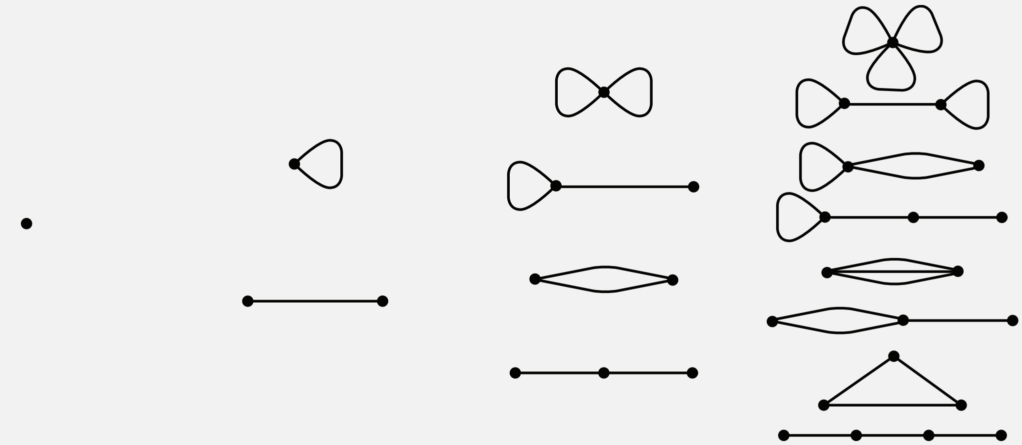
for any $Y \subseteq V_1 \uplus \dots V_2 \uplus \dots \uplus V_k$.

Exercise: Matroid Union, and Matroid duality

Exercise: Fully characterize $M \vee M^*$.

Matroids of three or fewer elements are graphic

- All matroids up to and including three elements (edges) are graphic.



(a) The only matroid with zero elements.
(b) The two one-element matroids.
(c) The four two-element matroids.
(d) The eight three-element matroids.

- This is a nice way to visualize matroids with very low ground set sizes. What about matroids that are low rank but with many elements?

Affine Matroids

- Given an $n \times m$ matrix with entries over some field $\mathbb{F}$, we say that a subset $S \subseteq \{1, \dots, m\}$ of indices (with corresponding column vectors $\{v_i : i \in S\}$, with $|S| = k$) is affinely dependent if $m \geq 1$ and there exists elements $\{a_1, \dots, a_k\} \in \mathbb{F}$, not all zero with $\sum_{i=1}^k a_i = 0$, such that $\sum_{i=1}^k a_i v_i = 0$.
- Otherwise, the set is called affinely independent.
- Concisely: points $\{v_1, v_2, \dots, v_k\}$ are affinely independent if $v_2 - v_1, v_3 - v_1, \dots, v_k - v_1$ are linearly independent.
- Example: in 2D, three collinear points are affinely dependent, three non-collear points are affinely independent, and ≥ 4 collinear or non-collinear points are affinely dependent.

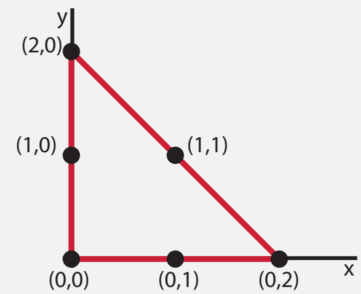
Proposition 9.5.1 (affine matroid)

Let ground set $E = \{1, \dots, m\}$ index column vectors of a matrix, and let $\mathcal{I}$ be the set of subsets X of E such that X indices affinely independent vectors. Then $(E, \mathcal{I})$ is a matroid.

Exercise: prove this.

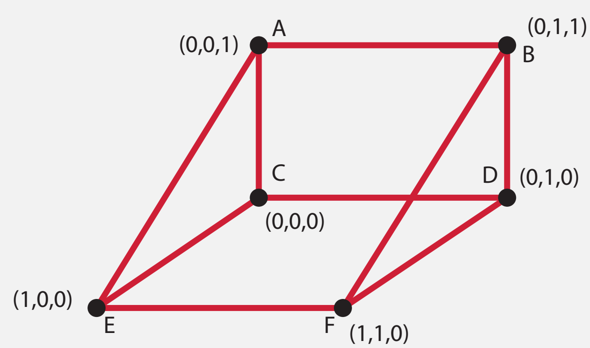
Euclidean Representation of Low-rank Matroids

- Consider the affine matroid with $n \times m = 2 \times 6$ matrix on the field $\mathbb{F} = \mathbb{R}$, and let the elements be $\{(0,0),(1,0),(2,0),(0,1),(0,2),(1,1)\}$.
- We can plot the points in $\mathbb{R}^2$ as on the right:
- Points have rank 1, lines have rank 2, planes have rank 3.
- Flats (points, lines, planes, etc.) have rank equal to one more than their geometric dimension.
- Any two points constitute a line, but lines with only two points are not drawn.
- Lines indicate collinear sets with ≥ 3 points, while any two points have rank 2.
- Dependent sets consist of all subsets with ≥ 4 elements (rank 3), or 3 collinear elements (rank 2). Any two points have rank 2.



Euclidean Representation of Low-rank Matroids

- As another example on the right, a rank 4 matroid



- All sets of 5 points are dependent. The only other sets of dependent points are coplanar ones of size 4. Namely:
  $\{(0,0,0),(0,1,0),(1,1,0),(1,0,0)\}$,
  $\{(0,0,0),(0,0,1),(0,1,1),(0,1,0)\}$, and
  $\{(0,0,1),(0,1,1),(1,1,0),(1,0,0)\}$.

Euclidean Representation of Low-rank Matroids

- In general, for a matroid $\mathcal{M}$ of rank $m+1$ with $m \leq 3$, then a subset X in a geometric representation in $\mathbb{R}^m$ is dependent if:
  1. $|X| \geq 2$ and the points are identical;
  2. $|X| \geq 3$ and the points are collinear;
  3. $|X| \geq 4$ and the points are coplanar; or
  4. $|X| \geq 5$ and the points are anywhere in space.
- When they exist, loops are represented in a geometry by a separate box indicating how many loops there are.
- Parallel elements, when they exist in a matroid, are indicated by a multiplicity next to a point.

Theorem 9.5.2

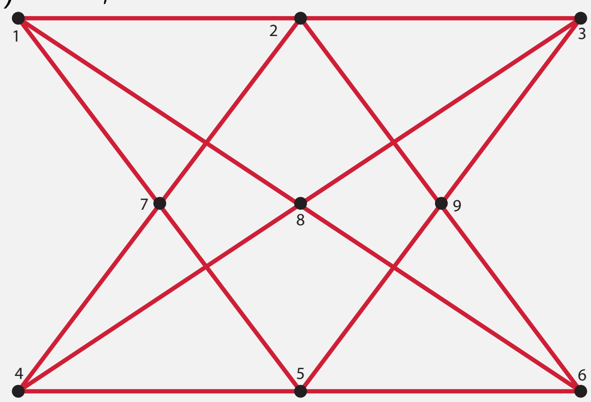
Any matroid of rank $m \leq 4$ can be represented by an affine matroid in $\mathbb{R}^{m-1}$.

Euclidean Rep. of Low-rank Matroids: Conditions

- rank-1 (resp. rank-2, rank-3) flats correspond to points (resp. lines, planes).
- a set of parallel points (could be size 1) does not touch another set of parallel points (could be size 1).
- every line contains at least two points (not dependent unless > 2).
- any two distinct points lie on a line (often not drawn when only two)
- every plane contains at least three non-collinear points (not dependent unless > 3)
- any three distinct non-collinear points lie on a plane
- If diagram has at most one plane, then any two distinct lines meet in at most one point.
- If diagram has more than one plane, then: 1) any two distinct planes meeting in more than two points do so in a line; 2) any two distinct lines meeting in a point do so in at most one point and lie in on a common plane; 3) any line not lying on a plane intersects it in at most one point.
- (see Oxley 2011 for more details).

Euclidean Representation of Low-rank Matroids

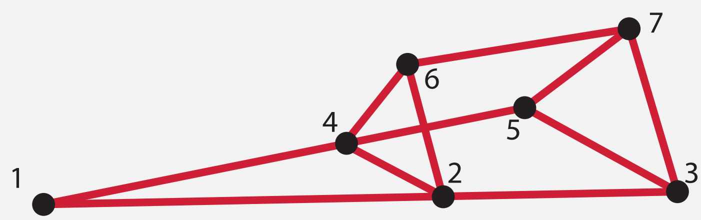
- Very useful for graphically depicting low-rank matrices but which still have rich structure. Also useful for answering questions.
- Example: Is there a matroid that is not representable (i.e., not linear for some field)? Yes, consider the matroid



- Called the non-Pappus matroid. Has rank three, but any matric matroid with the above dependencies would require that $\{7, 8, 9\}$ is dependent, hence requiring an additional line in the above.

Euclidean Representation of Low-rank Matroids: A test

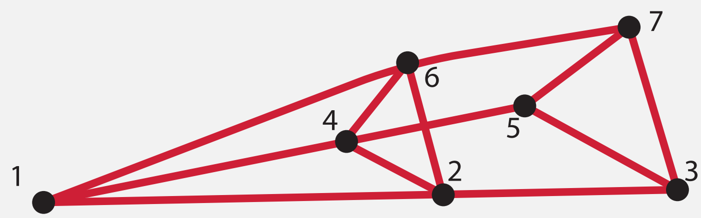
- Is this a matroid?



- Check rank's submodularity: Let $X = \{1, 2, 3, 6, 7\}$, $Y = \{1, 4, 5, 6, 7\}$. So $r(X) = 3$, and $r(Y) = 3$, and $r(X \cup Y) = 4$, so we must have, by submodularity, that $r(\{1, 6, 7\}) = r(X \cap Y) \leq r(X) + r(Y) - r(X \cup Y) = 2$.
- However, from the diagram, we have that since 1, 6, 7 are distinct non-collinear points, we have that $r(X \cap Y) = 3$

Euclidean Representation of Low-rank Matroids: A test

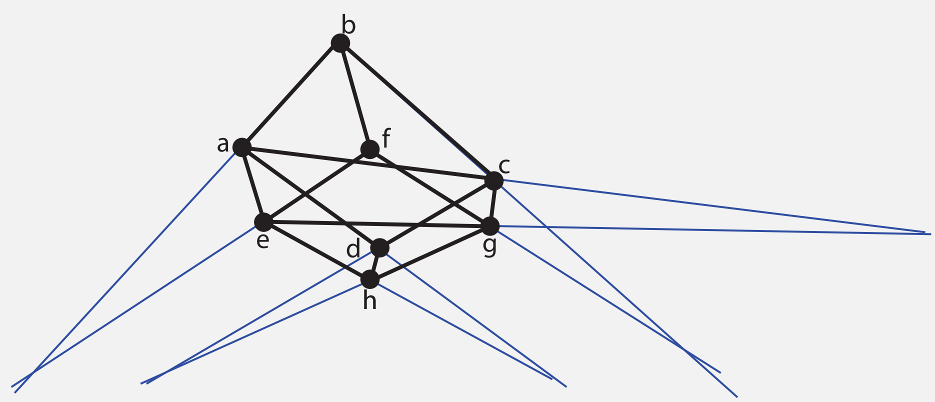
- Is this a matroid?



- If we extend the line from 6-7 to 1, then is it a matroid?
- Hence, not all 2D or 3D graphs of points and lines are matroids.

Matroid?

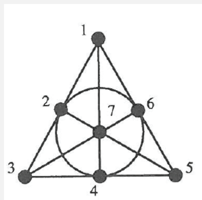
- Consider the following geometry on $|V| = 8$ points with $V = \{a, b, c, d, e, f, g, h\}$.

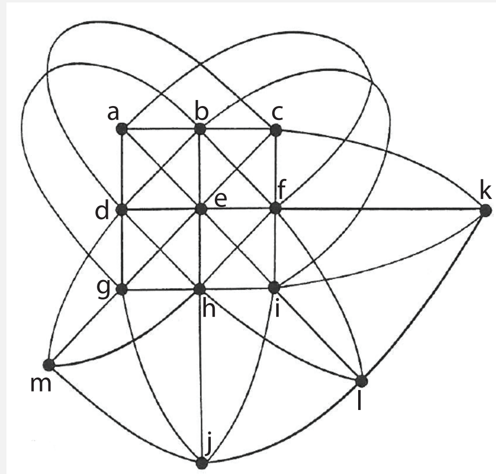


- Note, we are given that the points $\{b, d, h, f\}$ are not coplanar. However, the following sets of points are coplanar: $\{a, b, e, f\}$, $\{d, c, g, h\}$, $\{a, d, h, e\}$, $\{b, c, g, f\}$, $\{b, c, d, a\}$, $\{f, g, h, e\}$, and $\{a, c, g, e\}$.
- Exercise: Is this a matroid? Exercise: If so, is it representable?

Projective Geometries: Other Examples

- Other examples can be more complex, consider the following two matroids (from Oxley, 2011):





- Right: a matroid (and a 2D depiction of a geometry) over the field GF$(3) = \{0, 1, 2\}$ mod 3 and is "coordinatizable" in GF$(3)^3$.
- Hence, lines (in 2D) which are rank 2 sets may be curved; planes (in 3D) can be twisted.

Matroids, Representation and Equivalence: Summary

- Matroids with $|V| \leq 3$ are graphic.
- Matroids with $r(V) \leq 4$ can be geometrically represented in $\mathbb{R}^3$.
- Not all matroids are linear (i.e., matric) matroids.
- Matroids can be seen as related to projective geometries (and are sometimes called combinatorial geometries).
- Exists much research on different subclasses of matroids, and if/when they are contained in (or isomorphic to) each other.

Matroid Further Reading

- "Matroids: A Geometric Introduction", Gordon and McNulty, 2012.
- "The Coming of the Matroids", William Cunningham, 2012 (a nice history)
- Welsh, "Matroid Theory", 1975.
- Oxley, "Matroid Theory", 1992 (and 2011) (perhaps best "single source" on matroids right now).
- Crapo & Rota, "On the Foundations of Combinatorial Theory: Combinatorial Geometries", 1970 (while this is old, it is very readable).
- Lawler, "Combinatorial Optimization: Networks and Matroids", 1976.
- Schrijver, "Combinatorial Optimization", 2003

The greedy algorithm

- In combinatorial optimization, the greedy algorithm is often useful as a heuristic that can work quite well in practice.
- The goal is to choose a good subset of items, and the fundamental tenet of the greedy algorithm is to choose next whatever currently looks best, without the possibility of later recall or backtracking.
- Sometimes, this gives the optimal solution (we saw three greedy algorithms that can find the maximum weight spanning tree).
- Greedy is good since it can be made to run very fast $O(n \log n)$.
- Often, however, greedy is heuristic (it might work well in practice, but worst-case performance can be unboundedly poor).
- We will next see that the greedy algorithm working optimally is a defining property of a matroid, and is also a defining property of a polymatroid function.

Matroid and the greedy algorithm

- Let $(E, \mathcal{I})$ be an independence system, and we are given a non-negative modular weight function $w : E \to \mathbb{R}_+$.

Algorithm 1: The Matroid Greedy Algorithm

```
Set X ← ∅ ;
while ∃v ∈ E \ X s.t. X ∪ {v} ∈ I do
    v ∈ argmax {w(v) : v ∈ E \ X, X ∪ {v} ∈ I} ;
    X ← X ∪ {v} ;
```

- Same as sorting items by decreasing weight w, and then choosing items in that order that retain independence.

Theorem 9.6.1

Let $(E, \mathcal{I})$ be an independence system. Then the pair $(E, \mathcal{I})$ is a matroid if and only if for each weight function $w \in \mathcal{R}_+^E$, Algorithm 1 leads to a set $I \in \mathcal{I}$ of maximum weight $w(I)$.

Review from Lecture 6

- The next slide is from Lecture 6.

Matroids by bases

In general, besides independent sets and rank functions, there are other equivalent ways to characterize matroids.

Theorem 9.6.3 (Matroid (by bases))

Let E be a set and $\mathcal{B}$ be a nonempty collection of subsets of E. Then the following are equivalent.

1. *$\mathcal{B}$ is the collection of bases of a matroid;*
2. *if $B, B' \in \mathcal{B}$, and $x \in B' \setminus B$, then $B' - x + y \in \mathcal{B}$ for some $y \in B \setminus B'$.*
3. *If $B, B' \in \mathcal{B}$, and $x \in B' \setminus B$, then $B - y + x \in \mathcal{B}$ for some $y \in B \setminus B'$.*

Properties 2 and 3 are called "exchange properties."
Proof here is omitted but think about this for a moment in terms of linear spaces and matrices, and (alternatively) spanning trees.

Matroid and the greedy algorithm

proof of Theorem 9.6.1.

- Assume $(E, \mathcal{I})$ is a matroid and $w : E \to \mathcal{R}_+$ is given.
- Let $A = (a_1, a_2, \dots, a_r)$ be the solution returned by greedy, where $r = r(M)$ the rank of the matroid, and we order the elements as they were chosen (so $w(a_1) \geq w(a_2) \geq \cdots \geq w(a_r)$).
- A is a base of M, and let $B = (b_1, \dots, b_r)$ be <u>any</u> another base of M with elements also ordered decreasing by weight, so $w(b_1) \geq w(b_2) \geq \cdots \geq w(b_r)$.
- We next show that not only is $w(A) \geq w(B)$ but that $w(a_i) \geq w(b_i)$ for all i.

...

Matroid and the greedy algorithm

proof of Theorem 9.6.1.

- Assume otherwise, and let k be the first (smallest) integer such that $w(a_k) < w(b_k)$. Hence $w(a_j) \geq w(b_j)$ for $j < k$.
- Define independent sets $A_{k-1} = \{a_1, \dots, a_{k-1}\}$ and $B_k = \{b_1, \dots, b_k\}$.
- Since $|A_{k-1}| < |B_k|$, there exists a $b_i \in B_k \setminus A_{k-1}$ where $A_{k-1} \cup \{b_i\} \in \mathcal{I}$ for some $1 \leq i \leq k$.
- But $w(b_i) \geq w(b_k) > w(a_k)$, and so the greedy algorithm would have chosen b_i rather than a_k, contradicting what greedy does.

□

Matroid and the greedy algorithm

converse proof of Theorem 9.6.1.

- Given an independence system $(E, \mathcal{I})$, suppose the greedy algorithm leads to an independent set of max weight for every non-negative weight function. We'll show $(E, \mathcal{I})$ is a matroid.
- Emptyset containing and down monotonicity already holds (since we've started with an independence system).
- Let $I, J \in \mathcal{I}$ with $|I| < |J|$. Suppose to the contrary, that $I \cup \{z\} \notin \mathcal{I}$ for all $z \in J \setminus I$.
- Define the following modular weight function w on E, and define $k = |I|$.

$$w(v) = \begin{cases} k+2 & \text{if } v \in I, \\ k+1 & \text{if } v \in J \setminus I, \\ 0 & \text{if } v \in E \setminus (I \cup J) \end{cases} \tag{9.15}$$

...

Matroid and the greedy algorithm

converse proof of Theorem 9.6.1.

- Now greedy will, after k iterations, recover I, but it cannot choose any element in $J \setminus I$ by assumption. Thus, greedy chooses a set of weight $k(k+2)$.
- On the other hand, J has weight

$$w(J) \geq |J|(k+1) \geq (k+1)(k+1) > k(k+2) \tag{9.16}$$

  so J has strictly larger weight but is still independent, contradicting greedy's optimality.
- Therefore, there must be a $z \in J \setminus I$ such that $I \cup \{z\} \in \mathcal{I}$, and since I and J are arbitrary, $(E, \mathcal{I})$ must be a matroid.

□

Matroid and greedy

- As given, the theorem asked for a modular function $w \in \mathbb{R}_+^E$.
- This will not only return an independent set, but it will return a base if we keep going even if the weights are 0.
- If we don't want elements with weight 0, we can stop once (and if) the weight hits zero, thus giving us a maximum weight independent set.
- We don't need non-negativity, we can use any $w \in \mathbb{R}^E$ and keep going until we have a base.
- If we stop at a negative value, we'll once again get a maximum weight independent set.
- Exercise: what if we keep going until a base even if we encounter negative values?
- We can instead do as small as possible thus giving us a minimum weight independent set/base.

Summary of Important (for us) Matroid Definitions

Given an independence system, matroids are defined equivalently by any of the following:

- All maximally independent sets have the same size.
- A monotone non-decreasing submodular integral rank function with unit increments.
- The greedy algorithm achieves the maximum weight independent set for all weight functions.

Convex Polyhedra

- Convex polyhedra a rich topic, we will only draw what we need.

Definition 9.7.1

A subset $P \subseteq \mathbb{R}^E$ is a polyhedron if there exists an $m \times n$ matrix A and vector $b \in \mathbb{R}^m$ (for some $m \geq 0$) such that

$$P = \left\{x \in \mathbb{R}^E : Ax \leq b\right\} \tag{9.17}$$

- Thus, P is intersection of finitely many affine halfspaces, which are of the form $a_i x \leq b_i$ where a_i is a row vector and b_i a real scalar.

Convex Polytope

- A polytope is defined as follows

Definition 9.7.2

A subset $P \subseteq \mathbb{R}^E$ is a polytope if it is the convex hull of finitely many vectors in $\mathcal{R}^E$. That is, if $\exists$, $x_1, x_2, \dots, x_k \in \mathcal{R}^E$ such that for all $x \in P$, there exits $\{\lambda_i\}$ with $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0 \; \forall i$ with $x = \sum_i \lambda_i x_i$.

- We define the convex hull operator as follows:

$$\text{conv}(x_1, x_2, \dots, x_k) \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^{k} \lambda_i x_i : \forall i, \; \lambda_i \geq 0, \text{ and } \sum_i \lambda_i = 1 \right\} \tag{9.18}$$

Convex Polytope - key representation theorem

- A polytope can be defined in a number of ways, two of which include

Theorem 9.7.3

A subset $P \subseteq \mathbb{R}^E$ is a polytope iff it can be described in either of the following (equivalent) ways:

- *P is the convex hull of a finite set of points.*
- *If it is a bounded intersection of halfspaces, that is there exits matrix A and vector b such that*

$$P = \{x : Ax \leq b\} \tag{9.19}$$

- This result follows directly from results proven by Fourier, Motzkin, Farkas, and Carátheodory.

Linear Programming

Theorem 9.7.4 (weak duality)

Let A be a matrix and b and c vectors, then

$$\max\{c^\intercal x | Ax \leq b\} \leq \min\{y^\intercal b : y \geq 0, y^\intercal A = c^\intercal\} \tag{9.20}$$

Theorem 9.7.5 (strong duality)

Let A be a matrix and b and c vectors, then

$$\max\{c^\intercal x | Ax \leq b\} = \min\{y^\intercal b : y \geq 0, y^\intercal A = c^\intercal\} \tag{9.21}$$

Linear Programming duality forms

There are many ways to construct the dual. For example,

$$\max\{c^\intercal x | x \geq 0, Ax \leq b\} = \min\{y^\intercal b | y \geq 0, y^\intercal A \geq c^\intercal\} \tag{9.22}$$
$$\max\{c^\intercal x | x \geq 0, Ax = b\} = \min\{y^\intercal b | y^\intercal A \geq c^\intercal\} \tag{9.23}$$
$$\min\{c^\intercal x | x \geq 0, Ax \geq b\} = \max\{y^\intercal b | y \geq 0, y^\intercal A \leq c^\intercal\} \tag{9.24}$$
$$\min\{c^\intercal x | Ax \geq b\} = \max\{y^\intercal b | y \geq 0, y^\intercal A = c^\intercal\} \tag{9.25}$$

Linear Programming duality forms

How to form the dual in general? We quote V. Vazirani (2001)

> *Intuitively, why is [one set of equations] the dual of [another quite different set of equations]? In our experience, this is not the right question to be asked. As stated in Section 12.1, there is a purely mechanical procedure for obtaining the dual of a linear program. Once the dual is obtained, one can devise intuitive, and possibly physical meaningful, ways of thinking about it. Using this mechanical procedure, one can obtain the dual of a complex linear program in a fairly straightforward manner. Indeed, the LP-duality-based approach derives its wide applicability from this fact.*

Also see the text "Convex Optimization" by Boyd and Vandenberghe, chapter 5, for a great discussion on duality.

Vector, modular, incidence

- Recall, any vector $x \in \mathbb{R}^E$ can be seen as a normalized modular function, as for any $A \subseteq E$, we have

$$x(A) = \sum_{a \in A} x_a \tag{9.26}$$

- Given an $A \subseteq E$, define the the incidence vector $\mathbf{1}_A \in \{0,1\}^E$ on the unit hypercube as follows:

$$\mathbf{1}_A \stackrel{\text{def}}{=} \left\{ x \in \{0,1\}^E : x_i = 1 \text{ iff } i \in A \right\} \tag{9.27}$$

equivalently,

$$\mathbf{1}_A(j) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } j \in A \\ 0 & \text{if } j \notin A \end{cases} \tag{9.28}$$

Review from Lecture 6

The next slide is review from lecture 6.

Matroid

Slight modification (non unit increment) that is equivalent.

Definition 9.8.3 (Matroid-II)

A set system $(E, \mathcal{I})$ is a Matroid if

(I1') $\emptyset \in \mathcal{I}$

(I2') $\forall I \in \mathcal{I}, J \subset I \Rightarrow J \in \mathcal{I}$ (down-closed or subclusive)

(I3') $\forall I, J \in \mathcal{I}$, with $|I| > |J|$, then there exists $x \in I \setminus J$ such that $J \cup \{x\} \in \mathcal{I}$

Note (I1)=(I1'), (I2)=(I2'), and we get (I3)≡(I3') using induction.

Independence Polyhedra

- For each $I \in \mathcal{I}$ of a matroid $M = (E, \mathcal{I})$, we can form the incidence vector $\mathbf{1}_I$.
- Taking the convex hull, we get the independent set polytope, that is

$$P_{\text{ind. set}} = \text{conv}\left\{\bigcup_{I \in \mathcal{I}} \{\mathbf{1}_I\}\right\} \tag{9.29}$$

- Since $\{\mathbf{1}_I : I \in \mathcal{I}\} \subseteq P_{\text{ind. set}}$, we have $\max\{w(I) : I \in \mathcal{I}\} \leq \max\{w^\intercal x : x \in P_{\text{ind. set}}\}$.
- Now take the rank function r of M, and define the following polyhedron:

$$P_r^+ = \left\{x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E\right\} \tag{9.30}$$

- Now, take any $x \in P_{\text{ind. set}}$, then we have that $x \in P_r^+$ (or $P_{\text{ind. set}} \subseteq P_r^+$). We show this next.

$P_{\text{ind. set}} \subseteq P_r^+$

- If $x \in P_{\text{ind. set}}$, then

$$x = \sum_i \lambda_i \mathbf{1}_{I_i} \tag{9.31}$$

for some appropriate vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$.
- Clearly, for such x, $x \geq 0$.
- Now, for any $A \subseteq E$,

$$x(A) = x^\intercal \mathbf{1}_A = \sum_i \lambda_i {\mathbf{1}_{I_i}}^\intercal \mathbf{1}_A \tag{9.32}$$

$$\leq \sum_i \lambda_i \max_{j : I_j \subseteq A} \mathbf{1}_{I_j}(E) \tag{9.33}$$

$$= \max_{j : I_j \subseteq A} \mathbf{1}_{I_j}(E) = \max_{I \in \mathcal{I}} |A \cap I| \tag{9.34}$$

$$= r(A) \tag{9.35}$$

- Thus, $x \in P_r^+$ and hence $P_{\text{ind. set}} \subseteq P_r^+$.

Matroid Polyhedron in 2D

$$P_r^+ = \{x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E\} \quad (9.36)$$

- Consider this in two dimensions. We have equations of the form:

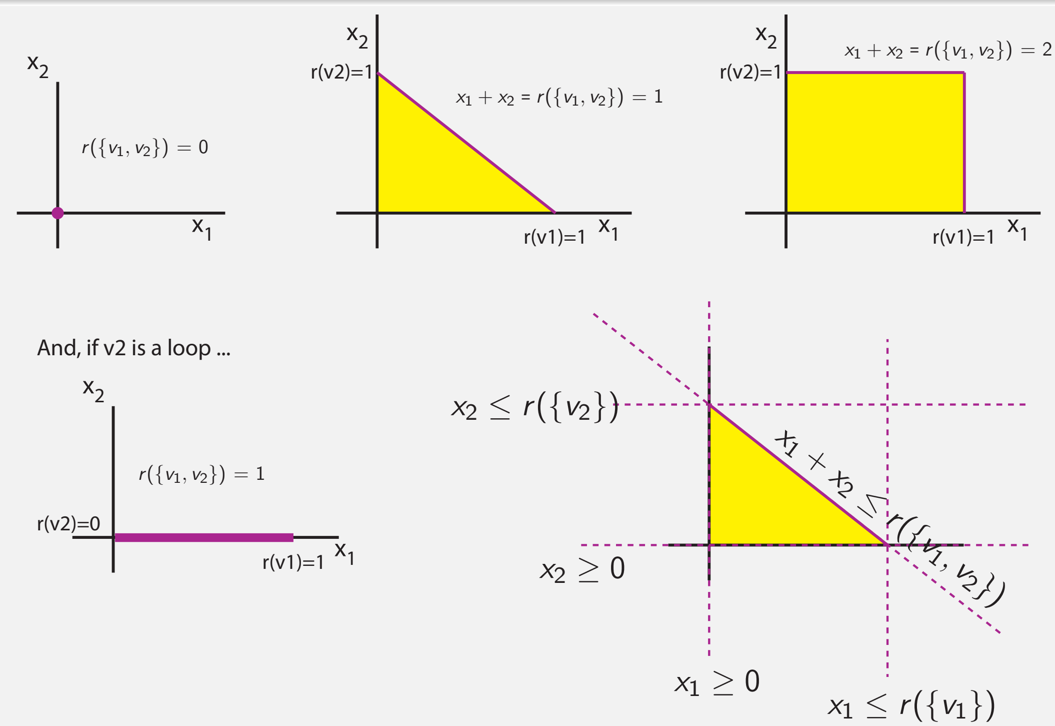
$$x_1 \geq 0 \text{ and } x_2 \geq 0 \quad (9.37)$$
$$x_1 \leq r(\{v_1\}) \quad (9.38)$$
$$x_2 \leq r(\{v_2\}) \quad (9.39)$$
$$x_1 + x_2 \leq r(\{v_1, v_2\}) \quad (9.40)$$

- Because r is submodular, we have

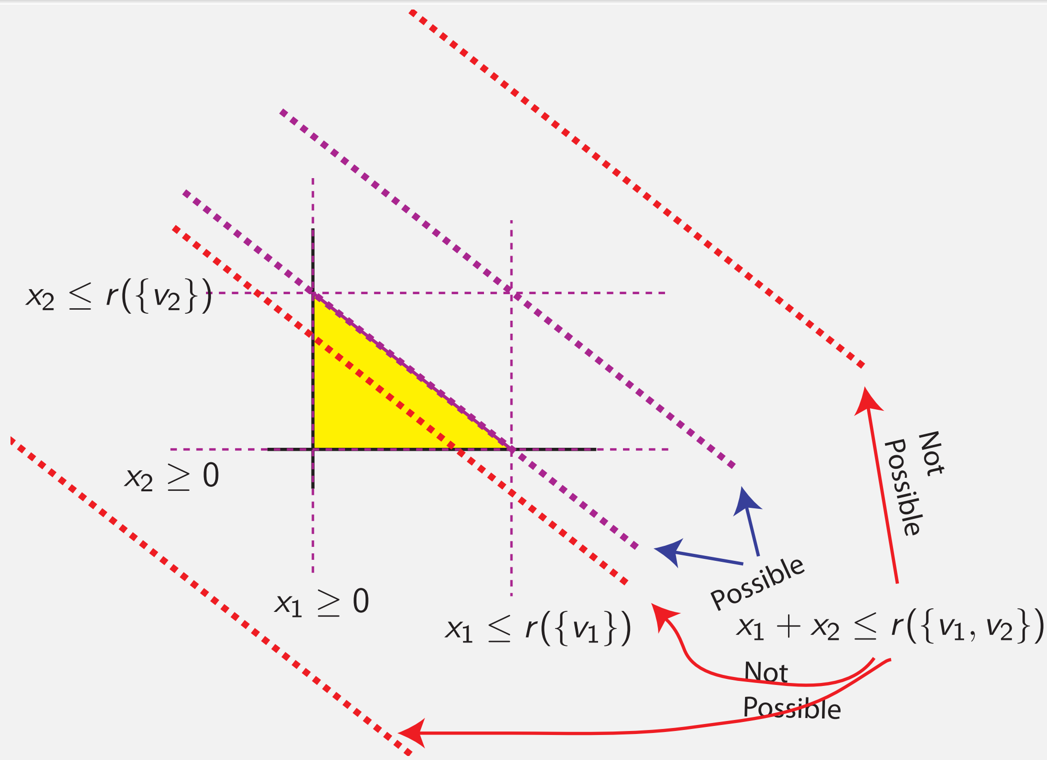
$$r(\{v_1\}) + r(\{v_2\}) \geq r(\{v_1, v_2\}) + r(\emptyset) \quad (9.41)$$

so since $r(\{v_1, v_2\}) \leq r(\{v_1\}) + r(\{v_2\})$, the last inequality is either touching (so inactive) or active.

Matroid Polyhedron in 2D



Matroid Polyhedron in 2D



Matroid Polyhedron in 3D

$$P_r^+ = \{x \in \mathbb{R}^E : x \ge 0, x(A) \le r(A), \forall A \subseteq E\} \tag{9.42}$$

- Consider this in three dimensions. We have equations of the form:

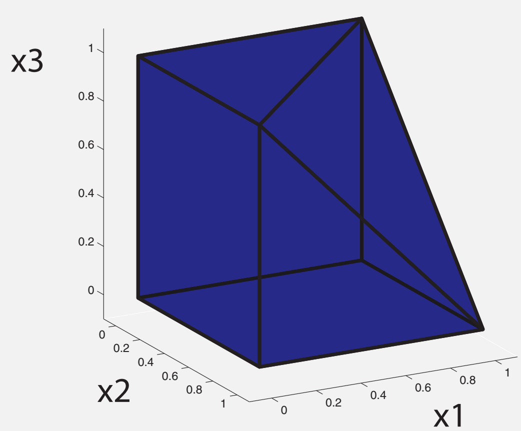
$$x_1 \ge 0 \text{ and } x_2 \ge 0 \text{ and } x_3 \ge 0 \tag{9.43}$$
$$x_1 \le r(\{v_1\}) \tag{9.44}$$
$$x_2 \le r(\{v_2\}) \tag{9.45}$$
$$x_3 \le r(\{v_3\}) \tag{9.46}$$
$$x_1 + x_2 \le r(\{v_1, v_2\}) \tag{9.47}$$
$$x_2 + x_3 \le r(\{v_2, v_3\}) \tag{9.48}$$
$$x_1 + x_3 \le r(\{v_1, v_3\}) \tag{9.49}$$
$$x_1 + x_2 + x_3 \le r(\{v_1, v_2, v_3\}) \tag{9.50}$$

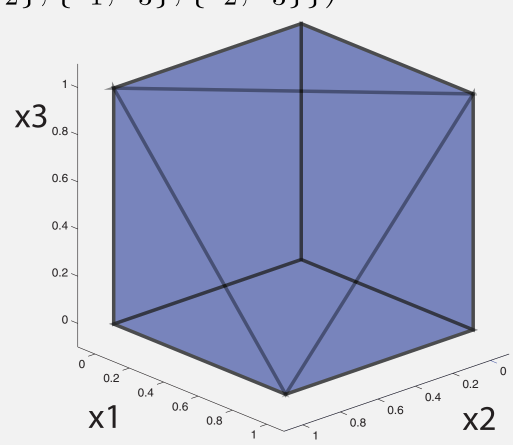
Matroid Polyhedron in 3D

- Consider the simple cycle matroid on a graph consisting of a 3-cycle, $G = (V, E)$ with matroid $M = (E, \mathcal{I})$ where $I \in \mathcal{I}$ is a forest.
- So any set of either one or two edges is independent, and has rank equal to cardinality.
- The set of three edges is dependent, and has rank 2.

Matroid Polyhedron in 3D

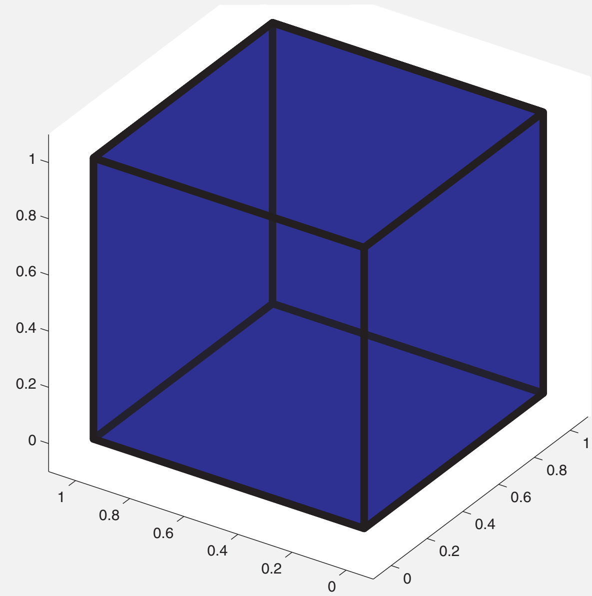
Two view of P_r^+ associated with a matroid $(\{e_1, e_2, e_3\}, \{\emptyset, \{e_1\}, \{e_2\}, \{e_3\}, \{e_1, e_2\}, \{e_1, e_3\}, \{e_2, e_3\}\})$.





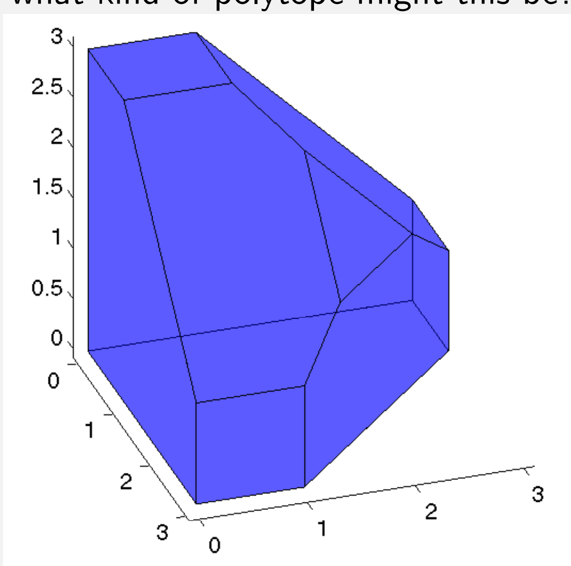
Matroid Polyhedron in 3D

P_r^+ associated with the "free" matroid in 3D.



Another Polytope in 3D

Thought question: what kind of polytope might this be?



Matroid Independence Polyhedron

- So recall from a moment ago, that we have that

$$P_{\text{ind. set}} = \text{conv}\left\{\cup_{I \in \mathcal{I}} \{\mathbf{1}_I\}\right\} \subseteq P_r^+ = \left\{x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E\right\} \tag{9.51}$$

- In fact, the two polyhedra are identical (and thus both are polytopes).
- We'll show this in the next few theorems.

Maximum weight independent set via greedy weighted rank

Theorem 9.8.1

Let $M = (V, \mathcal{I})$ be a matroid, with rank function r, then for any weight function $w \in \mathbb{R}_+^V$, there exists a chain of sets $U_1 \subset U_2 \subset \cdots \subset U_n \subseteq V$ such that

$$\max\left\{w(I) | I \in \mathcal{I}\right\} = \sum_{i=1}^{n} \lambda_i r(U_i) \tag{9.52}$$

where $\lambda_i \geq 0$ satisfy

$$w = \sum_{i=1}^{n} \lambda_i \mathbf{1}_{U_i} \tag{9.53}$$

Maximum weight independent set via weighted rank

Proof.

- Firstly, note that for any such $w \in \mathbb{R}^E$, we have

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = (w_1 - w_2)\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + (w_2 - w_3)\begin{pmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + (w_{n-1} - w_n)\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \end{pmatrix} + (w_n)\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix} \tag{9.54}$$

- If we can take w in decreasing order ($w_1 \geq w_2 \geq \cdots \geq w_n$), then each coefficient of the vectors is non-negative (except possibly the last one, w_n).

...

Maximum weight independent set via weighted rank

Proof.

- Now, again assuming $w \in \mathbb{R}^E_+$, order the elements of V as $(v_1, v_2, \ldots, v_n)$ such that $w(v_1) \geq w(v_2) \geq \cdots \geq w(v_n)$
- Define the sets U_i based on this order as follows, for $i = 0, \ldots, n$

$$U_i \stackrel{\text{def}}{=} \{v_1, v_2, \ldots, v_i\} \tag{9.55}$$

- Define the set I as those elements where the rank increases, i.e.:

$$I \stackrel{\text{def}}{=} \{v_i | r(U_i) > r(U_{i-1})\}. \tag{9.56}$$

Hence, given an i with $v_i \notin I$, $r(U_i) = r(U_{i-1})$.

- Therefore, I is the output of the greedy algorithm for $\max\{w(I)|I \in \mathcal{I}\}$. *since items v_i are ordered decreasing by $w(v_i)$, and we only choose the ones that increase the rank, which means they don't violate independence.*
- And therefore, I is a maximum weight independent set (even a base, actually).

...

Maximum weight independent set via weighted rank

Proof.

- Now, we define λ_i as follows

$$\lambda_i \stackrel{\text{def}}{=} w(v_i) - w(v_{i+1}) \text{ for } i = 1, \ldots, n-1 \tag{9.57}$$

$$\lambda_n \stackrel{\text{def}}{=} w(v_n) \tag{9.58}$$

- And the weight of the independent set $w(I)$ is given by

$$w(I) = \sum_{v \in I} w(v) = \sum_{i=1}^{n} w(v_i)\big(r(U_i) - r(U_{i-1})\big) \tag{9.59}$$

$$= w(v_n)r(U_n) + \sum_{i=1}^{n-1}\big(w(v_i) - w(v_{i+1})\big)r(U_i) = \sum_{i=1}^{n} \lambda_i r(U_i) \tag{9.60}$$

- Since we took $v_1, v_2, \ldots$ in decreasing order, for all i, and since $w \in \mathbb{R}_+^E$, we have $\lambda_i \geq 0$