Prof. Jeff Bilmes

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

April 23rd, 2018

Clockwise from top left: v
László Lovász
Jack Edmonds
Satoru Fujishige
George Nemhauser
Laurence Wolsey
András Frank
Lloyd Shapley
H. Narayanan
Robert Bixby
William Cunningham
William Tutte
Richard Rado
Alexander Schrijver
Garrett Birkhoff
Hassler Whitney
Richard Dedekind

Logistics
Review
Cumulative Outstanding Reading

- Read chapter 1 from Fujishige’s book.
- Read chapter 2 from Fujishige’s book.
Logistics

Announcements, Assignments, and Reminders

- If you have any questions about anything, please ask them via our discussion board (https://canvas.uw.edu/courses/1216339/discussion_topics).

Prof. Jeff Bilmes
EE563/Spring 2018/Submodularity - Lecture 9 - April 23rd, 2018
F3/58 (pg. 3/65)

Class Road Map - EE563

- L1(3/26): Motivation, Applications, & Basic Definitions
- L3(4/2): Info theory exs, more apps, definitions, graph/combinatorial examples
- L4(4/4): Graph and Combinatorial Examples, Matrix Rank, Examples and Properties, visualizations
- L6(4/11): Matroids, Matroid Examples, Matroid Rank, Partition/Laminar Matroids
- L7(4/16): Laminar Matroids, System of Distinct Reps, Transversals, Transversal Matroid, Matroid Representation, Dual Matroids
- L9(4/23): Polyhedra, Matroid Polytopes, Matroids → Polymatroids
- L10(4/25):
- L11(4/30):
- L12(5/2):
- L13(5/7):
- L14(5/9):
- L15(5/14):
- L16(5/16):
- L17(5/21):
- L18(5/23):
- L19(5/28): Memorial Day (holiday)

Last day of instruction, June 1st. Finals Week: June 2-8, 2018.
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

  1. Set $X \leftarrow \emptyset$ ;
  2. while $\exists v \in E \setminus X$ s.t. $X \cup \{v\} \in \mathcal{I}$ do
  3. \hspace{1em} $v \in \text{argmax} \{w(v) : v \in E \setminus X, \ X \cup \{v\} \in \mathcal{I}\}$ ;
  4. \hspace{1em} $X \leftarrow X \cup \{v\}$ ;

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

Theorem 9.2.8

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 above leads to a set $I \in \mathcal{I}$ of maximum weight $w(I)$. 

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 9.3.1**

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

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

(9.1)

- Thus, $P$ is intersection of finitely many ($\ell$) 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 9.3.2**

A subset $P \subseteq \mathbb{R}^E = \mathbb{R}^m$ is a *polytope* if it is the convex hull of finitely many vectors in $\mathbb{R}^E$. That is, if $\exists$, $x_1, x_2, \ldots, x_k \in \mathbb{R}^E$ such that for all $x \in P$, there exists $\{\lambda_i\}$ 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:

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

(9.2)

Convex Polytope - key representation theorem

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

**Theorem 9.3.3**

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

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

(9.3)

- This result follows directly from results proven by Fourier, Motzkin, Farkas, and Carathéodory.
Theorem 9.3.4 (weak duality)

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

$$\max \{ c^T x | Ax \leq b \} \leq \min \{ y^T b : y \geq 0, y^T A = c^T \}$$  \hspace{1cm} (9.4)$$

Theorem 9.3.5 (strong duality)

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

$$\max \{ c^T x | Ax \leq b \} = \min \{ y^T b : y \geq 0, y^T A = c^T \}$$  \hspace{1cm} (9.5)$$

Linear Programming duality forms

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

$$\max \{ c^T x | x \geq 0, Ax \leq b \} = \min \{ y^T b | y \geq 0, y^T A \geq c^T \}$$  \hspace{1cm} (9.6)$$

$$\max \{ c^T x | x \geq 0, Ax = b \} = \min \{ y^T b | y^T A \geq c^T \}$$  \hspace{1cm} (9.7)$$

$$\min \{ c^T x | x \geq 0, Ax \geq b \} = \max \{ y^T b | y \geq 0, y^T A \leq c^T \}$$  \hspace{1cm} (9.8)$$

$$\min \{ c^T x | Ax \geq b \} = \max \{ y^T b | y \geq 0, y^T A = c^T \}$$  \hspace{1cm} (9.9)$$
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 and easy mechanical ways to construct it.

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

$$x(A) = \sum_{a \in A} x_a$$  \hspace{1cm} (9.10)

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

$$1_A \overset{\text{def}}{=} \left\{ x \in \{0, 1\}^E : x_i = 1 \iff i \in A \right\}$$  \hspace{1cm} (9.11)

equivalently,

$$1_A(j) \overset{\text{def}}{=} \begin{cases} 1 & \text{if } j \in A \\ 0 & \text{if } j \notin A \end{cases}$$  \hspace{1cm} (9.12)
The next slide is review from lecture 6.

Slight modification (non unit increment) that is equivalent.

**Definition 9.4.3 (Matroid-II)**

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

1. $(I1')$ $\emptyset \in \mathcal{I}$
2. $(I2')$ $\forall I \in \mathcal{I}, J \subset I \Rightarrow J \in \mathcal{I}$ (down-closed or subclusive)
3. $(I3')$ $\forall I, J \in \mathcal{I},$ with $|I| > |J|,$ then there exists $x \in I \setminus J$ such that $J \cup \{x\} \in \mathcal{I}$

Note $(I1) \Rightarrow (I1'),$ $(I2) \Rightarrow (I2'),$ and we get $(I3) \equiv (I3')$ using induction.
Independence Polyhedra

- For each $I \in \mathcal{I}$ of a matroid $M = (E, \mathcal{I})$, we can form the incidence vector $1_I$.
- Taking the convex hull, we get the **independent set polytope**, that is
  
  $$P_{\text{ind. set}} = \text{conv}\left\{ \bigcup_{I \in \mathcal{I}} \{1_I\} \right\} \subseteq [0, 1]^E$$  
  \hspace{1cm} (9.13)

- Since $\{1_I : I \in \mathcal{I}\} \subseteq P_{\text{ind. set}} \subseteq P_r^+$, we have $\max \{w(I) : I \in \mathcal{I}\} \leq \max \{w^\top x : x \in P_{\text{ind. set}}\} \leq \max \{w^\top x : x \in P_r^+\}$

- Now take the rank function $r$ of $M$, and define the following polyhedron:
  
  $$P_r^+ \triangleq \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E \}$$  
  \hspace{1cm} (9.14)

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
  
  $$x = \sum_i \lambda_i 1_I_i$$  
  \hspace{1cm} (9.15)

  for some appropriate vector $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$.
- Clearly, for such $x$, $x \geq 0$.
- Now, for any $A \subseteq E$,
  
  $$x(A) = x^\top 1_A = \sum_i \lambda_i 1_I_i^\top 1_A$$  
  \hspace{1cm} (9.16)

  $$\leq \sum_i \lambda_i \max_{j : I_j \subseteq A} 1_{I_j}(E)$$  
  \hspace{1cm} (9.17)

  $$= \max_{j : I_j \subseteq A} 1_{I_j}(E) \max_{I \subseteq \mathcal{I}} |A \cap I|$$  
  \hspace{1cm} (9.18)

  $$= r(A)$$  
  \hspace{1cm} (9.19)

- Thus, $x \in P_r^+$ and hence $P_{\text{ind. set}} \subseteq P_r^+$.
\[ P^+_r = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E \} \quad (9.20) \]

- Consider this in two dimensions. We have equations of the form:
  \[ x_1 \geq 0 \quad \text{and} \quad x_2 \geq 0 \quad (9.21) \]
  \[ x_1 \leq r(\{v_1\}) \in \{0, 1\} \quad (9.22) \]
  \[ x_2 \leq r(\{v_2\}) \in \{0, 1\} \quad (9.23) \]
  \[ x_1 + x_2 \leq r(\{v_1, v_2\}) \in \{0, 1, 2\} \quad (9.24) \]

- Because \( r \) is submodular, we have
  \[ r(\{v_1\}) + r(\{v_2\}) \geq r(\{v_1, v_2\}) + r(\emptyset) \quad (9.25) \]

so since \( r(\{v_1, v_2\}) \leq r(\{v_1\}) + r(\{v_2\}) \), the last inequality is either touching \( (r(v_1, v_2) = r(v_1) + r(v_2), \text{inactive}) \) or active.
Consider this in three dimensions. We have equations of the form:

\[ P_{r}^{+} = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E \} \quad (9.26) \]

- \( x_1 \geq 0 \) and \( x_2 \geq 0 \) and \( x_3 \geq 0 \) \quad (9.27)
- \( x_1 \leq r(\{v_1\}) \) \quad (9.28)
- \( x_2 \leq r(\{v_2\}) \) \quad (9.29)
- \( x_3 \leq r(\{v_3\}) \) \quad (9.30)
- \( x_1 + x_2 \leq r(\{v_1, v_2\}) \) \quad (9.31)
- \( x_2 + x_3 \leq r(\{v_2, v_3\}) \) \quad (9.32)
- \( x_1 + x_3 \leq r(\{v_1, v_3\}) \) \quad (9.33)
- \( x_1 + x_2 + x_3 \leq r(\{v_1, v_2, v_3\}) \) \quad (9.34)
Consider the simple cycle matroid on a graph consisting of a 3-cycle, $G = (V, E)$ with matroid $M = (E, I)$ 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

$P_r^+$ associated with the “free” matroid in 3D.

Thought question: what kind of polytope might this be?
So recall from a moment ago, that we have that

\[
P_{\text{ind. set}} = \text{conv} \left\{ \bigcup_{I \in \mathcal{I}} \{1_I\} \right\}
\subseteq P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E \} \quad (9.35)
\]

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

**Theorem 9.4.1**

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

\[
\max \{ w(I) | I \in \mathcal{I} \} = \sum_{i=1}^{n} \lambda_i r(U_i) \quad (9.36)
\]

where \( \lambda_i \geq 0 \) satisfy

\[
w = \sum_{i=1}^{n} \lambda_i 1_{U_i} \quad (9.37)
\]
Maximum weight independent set via weighted rank

Proof.

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

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

(9.38)

- 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, $w_n$).

Proof.

- Now, again assuming $w \in \mathbb{R}^E_+$, order the elements of $V$ non-increasing by $w$ so $(v_1, v_2, \ldots, v_n)$ 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 $i = 0, \ldots, n$

$$
U_i \overset{\text{def}}{=} \{v_1, v_2, \ldots, v_i\}
$$

(9.39)

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

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

(9.40)

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 (can even be a base).
Maximum weight independent set via weighted rank

**Proof.**

- Now, we define $\lambda_i$ as follows
  \[
  0 \leq \lambda_i \overset{\text{def}}{=} w(v_i) - w(v_{i+1}) \quad \text{for } i = 1, \ldots, n - 1 \tag{9.41}
  \]
  \[
  \lambda_n \overset{\text{def}}{=} w(v_n) \tag{9.42}
  \]

- And the weight of the independent set $w(I)$ is given by
  \[
  w(I) = \sum_{v \in I} w(v) = \sum_{i=1}^{n} w(v_i) (r(U_i) - r(U_{i-1})) \tag{9.43}
  \]
  \[
  = w(v_n) r(U_n) + \sum_{i=1}^{n-1} (w(v_i) - w(v_{i+1})) r(U_i) = \sum_{i=1}^{n} \lambda_i r(U_i) \tag{9.44}
  \]

- Since we ordered $v_1, v_2, \ldots$ non-increasing by $w$, for all $i$, and since $w \in \mathbb{R}^E_+$, we have $\lambda_i \geq 0$.

---

**Linear Program LP**

Consider the linear programming primal problem

\[
\begin{align*}
\text{maximize} & \quad w^\top x \\
\text{subject to} & \quad x_v \geq 0 \quad (v \in V) \\
& \quad x(U) \leq r(U) \quad (\forall U \subseteq V) \tag{9.45}
\end{align*}
\]

And its convex dual (note $y \in \mathbb{R}^{2^n}_+$, $y_U$ is a scalar element within this exponentially big vector):

\[
\begin{align*}
\text{minimize} & \quad \sum_{U \subseteq V} y_U r(U), \\
\text{subject to} & \quad y_U \geq 0 \quad (\forall U \subseteq V) \\
& \quad \sum_{U \subseteq V} y_U 1_U \geq w \tag{9.46}
\end{align*}
\]

Thanks to strong duality, the solutions to these are equal to each other.
Consider the linear programming primal problem

\[
\begin{align*}
\text{maximize} & \quad w^\top x \\
n\text{s.t.} & \quad x_v \geq 0 \quad (v \in V) \\
& \quad x(U) \leq r(U) \quad (\forall U \subseteq V) \\
\end{align*}
\]

(9.47)

This is identical to the problem

\[
\text{max } w^\top x \text{ such that } x \in P_{r}^+ 
\]

(9.48)

where, again, \( P_{r}^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E \} \).

Therefore, since \( P_{\text{ind. set}} \subseteq P_{r}^+ \), the above problem can only have a larger solution. I.e.,

\[
\text{max } w^\top x \text{ s.t. } x \in P_{\text{ind. set}} \leq \text{max } w^\top x \text{ s.t. } x \in P_{r}^+. 
\]

(9.49)

Hence, we have the following relations:

\[
\text{max } \{ w(I) : I \in I \} \leq \text{max } \{ w^\top x : x \in P_{\text{ind. set}} \} \\
\leq \text{max } \{ w^\top x : x \in P_{r}^+ \} \\
\text{def } \alpha_{\text{min}} = \text{min } \left\{ \sum_{U \subseteq V} y_U r(U) : \forall U, y_U \geq 0; \sum_{U \subseteq V} y_U 1_U \geq w \right\} 
\]

(9.50)

(9.51)

(9.52)

Theorem 9.4.1 states that

\[
\text{max } \{ w(I) : I \in I \} = \sum_{i=1}^{n} \lambda_i r(U_i) 
\]

(9.53)

for the chain of \( U_i \)'s and \( \lambda_i \geq 0 \) that satisfies \( w = \sum_{i=1}^{n} \lambda_i 1_{U_i} \) (i.e., the r.h.s. of Eq. 9.53 is feasible w.r.t. the dual LP).

Therefore, we also have \( \text{max } \{ w(I) : I \in I \} \leq \alpha_{\text{min}} \) and

\[
\text{max } \{ w(I) : I \in I \} = \sum_{i=1}^{n} \lambda_i r(U_i) \geq \alpha_{\text{min}} 
\]

(9.54)
Polytope equivalence

- Hence, we have the following relations:
  \[
  \max \{ w(I) : I \in \mathcal{I} \} = \max \{ w^\top x : x \in P_{\text{ind. set}} \} \quad (9.50)
  \]
  \[
  = \max \{ w^\top x : x \in P_r^+ \} \quad (9.51)
  \]
  \[
  \overset{\text{def}}{=} \alpha_{\text{min}} = \min \left\{ \sum_{U \subseteq V} y_U r(U) : \forall U, y_U \geq 0; \sum_{U \subseteq V} y_U 1_U \geq w \right\} \quad (9.52)
  \]

- Therefore, all the inequalities above are equalities.
- And since \( w \in \mathbb{R}_E^+ \) is an arbitrary direction into the positive orthant, we see that \( P_r^+ = P_{\text{ind. set}} \).
- That is, we have just proven:

**Theorem 9.4.2**

\[
P_r^+ = P_{\text{ind. set}} \quad (9.55)
\]

Polytope Equivalence (Summarizing the above)

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

\[
P_{\text{ind. set}} = \text{conv} \{ \bigcup_{I \in \mathcal{I}} \{1_I\} \} \quad (9.56)
\]

- Now take the rank function \( r \) of \( M \), and define the following polytope:

\[
P_r^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq r(A), \forall A \subseteq E \} \quad (9.57)
\]

**Theorem 9.4.3**

\[
P_r^+ = P_{\text{ind. set}} \quad (9.58)
\]
Greedy solves a linear programming problem

- So we can describe the independence polytope of a matroid using the set of inequalities (an exponential number of them).
- In fact, considering equations starting at Eq 9.50, the LP problem with exponential number of constraints $\max \{ w^T x : x \in P_r^+ \}$ is identical to the maximum weight independent set problem in a matroid, and since greedy solves the latter problem exactly, we have also proven:

**Theorem 9.4.4**

The LP problem $\max \{ w^T x : x \in P_r^+ \}$ can be solved exactly using the greedy algorithm.

Note that this LP problem has an exponential number of constraints (since $P_r^+$ is described as the intersection of an exponential number of half spaces).
- This means that if LP problems have certain structure, they can be solved much easier than immediately implied by the equations.

**Base Polytope Equivalence**

- Consider convex hull of indicator vectors just of the bases of a matroid, rather than all of the independent sets.
- Consider a polytope defined by the following constraints:

\[
\begin{align*}
x &\geq 0 \\
x(A) &\leq r(A) \quad \forall A \subseteq V \\
x(V) &= r(V)
\end{align*}
\]

(9.59) (9.60) (9.61)

- Note the third requirement, $x(V) = r(V)$.
- By essentially the same argument as above (Exercise:), we can shown that the convex hull of the incidence vectors of the bases of a matroid is a polytope that can be described by Eq. 9.59- 9.61 above.
- What does this look like?
Spanning set polytope

- Recall, a set $A$ is spanning in a matroid $M = (E, \mathcal{I})$ if $r(A) = r(E)$.
- Consider convex hull of incidence vectors of spanning sets of a matroid $M$, and call this $P_{\text{spanning}}(M)$.

**Theorem 9.4.5**

The spanning set polytope is determined by the following equations:

\[
0 \leq x_e \leq 1 \quad \text{for } e \in E \quad (9.62)
\]

\[
x(A) \geq r(E) - r(E \setminus A) \quad \text{for } A \subseteq E \quad (9.63)
\]

- Example of spanning set polytope in 2D.

\[
x_1 + x_2 = r(\{v_1, v_2\}) = 1
\]

**Proof.**

- Recall that any $A$ is spanning in $M$ iff $E \setminus A$ is independent in $M^*$ (the dual matroid).
- For any $x \in \mathbb{R}^E$, we have that

\[
x \in P_{\text{spanning}}(M) \iff 1 - x \in P_{\text{ind. set}}(M^*) \quad (9.64)
\]

as we show next . . .
Polyhedra

Matroid Polytopes

Matroids → Polymatroids

Spanning set polytope

... proof continued.

This follows since if \( x \in P_{\text{spanning}}(M) \), we can represent \( x \) as a convex combination:
\[
x = \sum_i \lambda_i 1_{A_i} \tag{9.65}
\]

where \( A_i \) is spanning in \( M \).

Consider
\[
1 - x = 1_E - x = 1_E - \sum_i \lambda_i 1_{A_i} = \sum_i \lambda_i 1_{E \setminus A_i}, \tag{9.66}
\]

which follows since \( \sum_i \lambda_i 1 = 1_E \), so \( 1 - x \) is a convex combination of independent sets in \( M^* \) and so \( 1 - x \in P_{\text{ind. set}}(M^*) \).

...
We’ve been discussing results about matroids (independence polytope, etc.).

By now, it is clear that matroid rank functions are special cases of submodular functions. We ultimately will be reviewing submodular function minimization procedures, but in some cases it is worth showing a result for a general submodular function first.

Henceforth, we will skip between submodular functions and matroids, each lecture talking less about matroids specifically and taking more about submodular functions more generally ...

Maximal points in a set

Regarding sets, a subset \( X \) of \( S \) is a maximal subset of \( S \) possessing a given property \( \mathcal{P} \) if \( X \) possesses property \( \mathcal{P} \) and no set properly containing \( X \) (i.e., any \( X' \supset X \) with \( X' \setminus X \subseteq V \setminus X \)) possesses \( \mathcal{P} \).

Given any compact (essentially closed & bounded) set \( P \subseteq \mathbb{R}^E \), we say that a vector \( x \) is maximal within \( P \) if it is the case that for any \( \epsilon > 0 \), and for all directions \( e \in E \), we have that

\[
x + \epsilon e \notin P
\]

Examples of maximal regions (in red)
Maximal points in a set

- Regarding sets, a subset $X$ of $S$ is a **maximal** subset of $S$ possessing a given property $\mathcal{P}$ if $X$ possesses property $\mathcal{P}$ and no set properly containing $X$ (i.e., any $X' \supset X$ with $X' \setminus X \subseteq V \setminus X$) possesses $\mathcal{P}$.
- Given any compact (essentially closed & bounded) set $P \subseteq \mathbb{R}^E$, we say that a vector $x$ is maximal within $P$ if it is the case that for any $\epsilon > 0$, and for all directions $e \in E$, we have that
  \[ x + \epsilon 1_e \notin P \]  
  (9.71)
- Examples of non-maximal regions (in green)

Review from Lecture 6

- The next slide comes from Lecture 6.
Matroids, independent sets, and bases

- **Independent sets**: Given a matroid $M = (E, \mathcal{I})$, a subset $A \subseteq E$ is called **independent** if $A \in \mathcal{I}$ and otherwise $A$ is called **dependent**.

- **A base of $U \subseteq E$**: For $U \subseteq E$, a subset $B \subseteq U$ is called a base of $U$ if $B$ is inclusionwise maximally independent subset of $U$. That is, $B \in \mathcal{I}$ and there is no $Z \in \mathcal{I}$ with $B \subset Z \subseteq U$.

- **A base of a matroid**: If $U = E$, then a “base of $E$” is just called a base of the matroid $M$ (this corresponds to a basis in a linear space, or a spanning forest in a graph, or a spanning tree in a connected graph).

---

**Definition 9.5.1 (subvector)**

$y$ is a **subvector** of $x$ if $y \leq x$ (meaning $y(e) \leq x(e)$ for all $e \in E$).

**Definition 9.5.2 ($P$-basis)**

Given a compact set $P \subseteq \mathbb{R}^E_+$, for any $x \in \mathbb{R}^E_+$, a subvector $y$ of $x$ is called a **$P$-basis** of $x$ if $y$ maximal in $P$. In other words, $y$ is a $P$-basis of $x$ if $y$ is a maximal $P$-contained subvector of $x$.

Here, by $y$ being “maximal”, we mean that there exists no $z > y$ (more precisely, no $z \geq y + \epsilon 1_e$ for some $e \in E$ and $\epsilon > 0$) having the properties of $y$ (the properties of $y$ being: in $P$, and a subvector of $x$).

In still other words: $y$ is a $P$-basis of $x$ if:

1. $y \leq x$ (y is a subvector of $x$); and
2. $y \in P$ and $y + \epsilon 1_e \notin P$ for all $e \in E$ where $y(e) < x(e)$ and $\forall \epsilon > 0$ ($y$ is maximal $P$-contained).
A vector form of rank

- Recall the definition of rank from a matroid \( M = (E, \mathcal{I}) \).
  \[
  \text{rank}(A) = \max \{|I| : I \subseteq A, I \in \mathcal{I}\} = \max_{I \in \mathcal{I}} |A \cap I| \quad (9.72)
  \]

- **vector rank**: Given a compact set \( P \subseteq \mathbb{R}^E_+ \), we can define a form of “vector rank” relative to this \( P \) in the following way: Given an \( x \in \mathbb{R}^E \), we define the vector rank, relative to \( P \), as:
  \[
  \text{rank}(x) = \max \{y(E) : y \leq x, y \in P\} = \max_{y \in P} (x \land y)(E) \quad (9.73)
  \]
  where \( y \leq x \) is componentwise inequality \((y_i \leq x_i, \forall i)\), and where \((x \land y) \in \mathbb{R}^E_+\) has \((x \land y)(i) = \min\{x(i), y(i)\}\).

- If \( B_x \) is the set of \( P \)-bases of \( x \), then \( \text{rank}(x) = \max_{y \in B_x} y(E) \).
- If \( x \in P \), then \( \text{rank}(x) = x(E) \) (\( x \) is its own unique self \( P \)-basis).
- If \( x_{\text{min}} = \min_{x \in P} x(E) \), and \( x \leq x_{\text{min}} \) what then? \(-\infty\)?

In general, might be hard to compute and/or have ill-defined properties. Next, we look at an object that restrains and cultivates this form of rank.

Polymatroidal polyhedron (or a “polymatroid”)

**Definition 9.5.3 (polymatroid)**

A **polymatroid** is a compact set \( P \subseteq \mathbb{R}^E_+ \) satisfying

1. \( 0 \in P \)
2. If \( y \leq x \in P \) then \( y \in P \) (called **down monotone**).
3. For every \( x \in \mathbb{R}^E_+ \), any maximal vector \( y \in P \) with \( y \leq x \) (i.e., any \( P \)-basis of \( x \)), has the same component sum \( y(E) \)

- Condition 3 restated: That is for any two distinct maximal vectors \( y^1, y^2 \in P \), with \( y^1 \leq x \) & \( y^2 \leq x \), with \( y^1 \neq y^2 \), we must have \( y^1(E) = y^2(E) \).
- Condition 3 restated (again): For every vector \( x \in \mathbb{R}^E_+ \), every maximal independent (i.e., \( \in P \)) subvector \( y \) of \( x \) has the same component sum \( y(E) = \text{rank}(x) \).
- Condition 3 restated (yet again): All \( P \)-bases of \( x \) have the same component sum.
Definition 9.5.3 (polymatroid)

A **polymatroid** is a compact set $P \subseteq \mathbb{R}^E_+$ satisfying:

1. $0 \in P$
2. If $y \leq x \in P$ then $y \in P$ (called **down monotone**).
3. For every $x \in \mathbb{R}^E_+$, any maximal vector $y \in P$ with $y \leq x$ (i.e., any $P$-basis of $x$), has the same component sum $y(E)$.

- Vectors within $P$ (i.e., any $y \in P$) are called **independent**, and any vector outside of $P$ is called **dependent**.
- Since all $P$-bases of $x$ have the same component sum, if $B_x$ is the set of $P$-bases of $x$, than $\text{rank}(x) = y(E)$ for any $y \in B_x$.

Matroid and Polymatroid: side-by-side

A Matroid is:

1. a set system $(E, I)$
2. empty-set containing $\emptyset \in I$
3. down closed, $\emptyset \subseteq I' \subseteq I \in I \Rightarrow I' \in I$.
4. any maximal set $I$ in $I$, bounded by another set $A$, has the same matroid rank (any maximal independent subset $I \subseteq A$ has same size $|I|$).

A Polymatroid is:

1. a compact set $P \subseteq \mathbb{R}^E_+$
2. zero containing, $0 \in P$
3. down monotone, $0 \leq y \leq x \in P \Rightarrow y \in P$
4. any maximal vector $y$ in $P$, bounded by another vector $x$, has the same vector rank (any maximal independent subvector $y \leq x$ has same sum $y(E)$).
Polymatroidal polyhedron (or a “polymatroid”)

Left: ∃ multiple maximal \( y \leq x \) Right: ∃ only one maximal \( y \leq x \),

- Polymatroid condition here: \( \forall \) maximal \( y \in P \), with \( y \leq x \) (which here means \( y_1 \leq x_1 \) and \( y_2 \leq x_2 \)), we just have \( y(E) = y_1 + y_2 = \text{const} \).
- On the left, we see there are multiple possible maximal \( y \in P \) such that \( y \leq x \). Each such \( y \) must have the same value \( y(E) \).
- On the right, there is only one maximal \( y \in P \). Since there is only one, the condition on the same value of \( y(E), \forall y \) is vacuous.

∃ only one maximal \( y \leq x \).

- If \( x \in P \) already, then \( x \) is its own \( P \)-basis, i.e., it is a self \( P \)-basis.
- In a matroid, a base of \( A \) is the maximally contained independent set. If \( A \) is already independent, then \( A \) is a self-base of \( A \) (as we saw in previous Lectures)
Polymatroid as well? no

Left and right: $\exists$ multiple maximal $y \leq x$ as indicated.

- On the left, we see there are multiple possible maximal such $y \in P$ that are $y \leq x$. Each such $y$ must have the same value $y(E)$, but since the equation for the curve is $y_1^2 + y_2^2 = \text{const.} \neq y_1 + y_2$, we see this is not a polymatroid.

- On the right, we have a similar situation, just the set of potential values that must have the $y(E)$ condition changes, but the values of course are still not constant.

Other examples: Polymatroid or not?
It appears that we have five possible forms of polymatroid in 2D, when neither of the elements \( \{v_1, v_2\} \) are self-dependent.

1. On the left: full dependence between \( v_1 \) and \( v_2 \)
2. Next: full independence between \( v_1 \) and \( v_2 \)
3. Next: partial independence between \( v_1 \) and \( v_2 \)
4. Right two: other forms of partial independence between \( v_1 \) and \( v_2 \)
   - The \( P \)-bases (or single \( P \)-base in the middle case) are as indicated.
   - Independent vectors are those within or on the boundary of the polytope. Dependent vectors are exterior to the polytope.
   - The set of \( P \)-bases for a polytope is called the base polytope.

Polymatroidal polyhedron (or a “polymatroid”)

- Note that if \( x \) contains any zeros (i.e., suppose that \( x \in \mathbb{R}_+^E \) has \( E \setminus S \) s.t. \( x(E \setminus S) = 0 \), so \( S \) indicates the non-zero elements, or \( S = \text{supp}(x) \)), then this also forces \( y(E \setminus S) = 0 \), so that \( y(E) = y(S) \). This is true either for \( x \in P \) or \( x \notin P \).
- Therefore, in this case, it is the non-zero elements of \( x \), corresponding to elements \( S \) (i.e., the support \( \text{supp}(x) \) of \( x \)), determine the common component sum.
- For the case of either \( x \notin P \) or right at the boundary of \( P \), we might give a “name” to this component sum, lets say \( f(S) \) for any given set \( S \) of non-zero elements of \( x \). We could name \( \text{rank}(\frac{1}{\epsilon} 1_S) \triangleq f(S) \) for \( \epsilon \) small enough. What kind of function might \( f \) be?
Definition 9.5.4

A **polymatroid function** is a real-valued function $f$ defined on subsets of $E$ which is normalized, non-decreasing, and submodular. That is we have

1. $f(\emptyset) = 0$ (normalized)
2. $f(A) \leq f(B)$ for any $A \subseteq B \subseteq E$ (monotone non-decreasing)
3. $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$ for any $A, B \subseteq E$ (submodular)

We can define the polyhedron $P_f^+$ associated with a polymatroid function as follows

$$P_f^+ = \{ y \in \mathbb{R}^E_+ : y(A) \leq f(A) \text{ for all } A \subseteq E \} \quad (9.74)$$

$$= \{ y \in \mathbb{R}^E : y \geq 0, y(A) \leq f(A) \text{ for all } A \subseteq E \} \quad (9.75)$$

Associated polyhedron with a polymatroid function

$$P_f^+ = \{ x \in \mathbb{R}^E : x \geq 0, x(A) \leq f(A), \forall A \subseteq E \} \quad (9.76)$$

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

  $x_1 \geq 0$ and $x_2 \geq 0$ and $x_3 \geq 0$ \quad (9.77)

  $x_1 \leq f(\{v_1\})$ \quad (9.78)

  $x_2 \leq f(\{v_2\})$ \quad (9.79)

  $x_3 \leq f(\{v_3\})$ \quad (9.80)

  $x_1 + x_2 \leq f(\{v_1, v_2\})$ \quad (9.81)

  $x_2 + x_3 \leq f(\{v_2, v_3\})$ \quad (9.82)

  $x_1 + x_3 \leq f(\{v_1, v_3\})$ \quad (9.83)

  $x_1 + x_2 + x_3 \leq f(\{v_1, v_2, v_3\})$ \quad (9.84)
Consider the asymmetric graph cut function on the simple chain graph \( v_1 - v_2 - v_3 \). That is, 
\[
f(S) = |\{(v, s) \in E(G) : v \in V, s \in S\}|
\]
is count of any edges within \( S \) or between \( S \) and \( V \setminus S \), so that 
\[
\delta(S) = f(S) + f(V \setminus S) - f(V)
\]
is the standard graph cut.

Observe: \( P_f^+ \) (at two views):

Consider: 
\[
\begin{align*}
f(\emptyset) &= 0, \\ f(\{v_1\}) &= 1.5, \\ f(\{v_2\}) &= 2, \\ f(\{v_1, v_2\}) &= 2.5, \\ f(\{v_3\}) &= 3, \\ f(\{v_3, v_1\}) &= 3.5, \\ f(\{v_3, v_2\}) &= 4, \\ f(\{v_3, v_2, v_1\}) &= 4.3.
\end{align*}
\]

Observe: \( P_f^+ \) (at two views):
**Polyhedra**

**Matroid Polytopes**

**Matroids**

→

**Polymatroids**

Associated polyhedron with a polymatroid function

- Consider modular function $w : V \rightarrow \mathbb{R}_+$ as $w = (1, 1.5, 2)^T$, and then the submodular function $f(S) = \sqrt{w(S)}$.
- Observe: $P_f^+$ (at two views):

![Polyhedron Diagram](image)

- which axis is which?

Prof. Jeff Bilmes  
EE563/Spring 2018/Submodularity - Lecture 9 - April 23rd, 2018  
F57/68 (pg.63/65)

**Associated polytope with a non-submodular function**

- Consider function on integers: $g(0) = 0$, $g(1) = 3$, $g(2) = 4$, and $g(3) = 5.5$. Is $f(S) = g(|S|)$ submodular? $f(S) = g(|S|)$ is not submodular since $f(\{e_1, e_3\}) + f(\{e_1, e_2\}) = 4 + 4 = 8$ but $f(\{e_1, e_2, e_3\}) + f(\{e_1\}) = 5.5 + 3 = 8.5$. Alternatively, consider concavity violation, $1 = g(1 + 1) - g(1) < g(2 + 1) - g(2) = 1.5$.
- Observe: $P_f^+$ (at two views), maximal independent subvectors not constant rank, hence **not** a polymatroid.

![Polytope Diagram](image)
A polymatroid vs. a polymatroid function’s polyhedron

- Summarizing the above, we have:
  - Given a polymatroid function $f$, its associated polytope is given as
    \[ P_f^+ = \{ y \in \mathbb{R}_+^E : y(A) \leq f(A) \text{ for all } A \subseteq E \} \] (9.85)
  - We also have the definition of a polymatroidal polytope $P$ (compact subset, zero containing, down-monotone, and $\forall x$ any maximal independent subvector $y \leq x$ has same component sum $y(E)$).

- Is there any relationship between these two polytopes?
- In the next theorem, we show that any $P_f^+$-basis has the same component sum, when $f$ is a polymatroid function, and $P_f^+$ satisfies the other properties so that $P_f^+$ is a polymatroid.