Logistics

Announcements, Assignments, and Reminders

- Please do use our discussion board (https://catalyst.uw.edu/gopost/board/bilmes/29948/) for all questions, comments, so that all will benefit from them being answered.
- I’ll likely be gone the week of Dec 3rd-7th. Hence, we’ll most likely have two extra lectures the following week (during finals week). More details TBA.
Cumulative Outstanding Reading

- Read Tom McCormick's overview paper on SFM http://people.commerce.ubc.ca/faculty/mccormick/sfmchap8a.pdf
- Read chapters 1 - 3 from Fujishige book.
- Read over lecture slides, all posted on our web page (http://j.ee.washington.edu/~bilmes/classes/ee596a_fall_2012/).
- See the summary slide at the end for lectures for additional reading sources.
The “integral” in the Choquet integral

- Thought of as an integral over $\mathbb{R}$ of a piece-wise constant function.
- First note, assuming $E$ is ordered according to descending $w$, so that $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_{m-1}) \geq w(e_m)$, then $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e \geq w_{e_i}\}$.
- For any $w_{e_i} > \alpha \geq w_{e_{i+1}}$ we also have $E_i = \{e_1, e_2, \ldots, e_i\} = \{e \in E : w_e > \alpha\}$.
- Consider segmenting the real-axis at boundary points $w_{e_i}$, right most is $w_{e_1}$.

\[
\begin{array}{ccccccc}
& & & & & w(e_m) & w(e_{m-1}) & \cdots & w(e_5) & w(e_4) & w(e_3) & w(e_2) & w(e_1)
\end{array}
\]

- A function can be defined on a segment of $\mathbb{R}$, namely $w_{e_i} > \alpha \geq w_{e_{i+1}}$. This function $F_i : [w_{e_{i+1}}, w_{e_i}) \rightarrow \mathbb{R}$ is defined as

\[
F_i(\alpha) = f(\{e \in E : w_e > \alpha\}) = f(E_i) \quad (14.11)
\]

We can generalize this to multiple segments of $\mathbb{R}$ (for now, take $w \in \mathbb{R}^E_+$. The piecewise-constant function is defined as:

\[
F(\alpha) = \begin{cases} 
    f(E) & \text{if } 0 \leq \alpha \leq w_m \\
    f(\{e \in E : w_e > \alpha\}) & \text{if } w_{e_{i+1}} \leq \alpha < w_{e_i}, \ i \in \{1, \ldots, m-1\} \\
    0 & \text{if } w_1 < \alpha
\end{cases}
\]

- Visualizing a piecewise constant function, where the constant values are given by $f$ evaluated on $E_i$ for each $i$

\[
\begin{array}{ccccccc}
f(E) & f(\{e \in E : w_e > \alpha\}) & f(E \setminus \{e_m, e_{m-1}\}) & f(\{e \in E : w_e > \alpha\}) & f(\{e \in E : w_e > \alpha\}) & f(E \setminus \{e_{m-1}\}) & f(E \setminus \{e_{m-2}\}) & f(E \setminus \{e_{m-3}\}) & f(E \setminus \{e_{m-4}\}) & f(E \setminus \{e_{m-5}\})
\end{array}
\]

Note, what is depicted may be a game but not a capacity.
The “integral” in the Choquet integral

- Now consider the integral, with \( w \in \mathbb{R}_+^E \), and normalized \( f \) so that \( f(\emptyset) = 0 \). Recall \( w_{m+1} \overset{\text{def}}{=} 0 \).

\[
\tilde{f}(w) \overset{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \quad (14.12)
\]

\[
= \int_0^\infty f(\{ e \in E : w_e > \alpha \}) d\alpha \quad (14.13)
\]

\[
= \int_{w_{m+1}}^\infty f(\{ e \in E : w_e > \alpha \}) d\alpha \quad (14.14)
\]

\[
= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(\{ e \in E : w_e > \alpha \}) d\alpha \quad (14.15)
\]

\[
= \sum_{i=1}^m \int_{w_{i+1}}^{w_i} f(E_i) d\alpha = \sum_{i=1}^m f(E_i)(w_i - w_{i+1}) \quad (14.16)
\]

- But we saw before that \( \sum_{i=1}^m f(E_i)(w_i - w_{i+1}) \) is just the Lovász extension of a function \( f \).

- Thus, we have the following definition:

**Definition 14.2.2**

Given \( w \in \mathbb{R}_+^E \), the Lovász extension (equivalently Choquet integral) may be defined as follows:

\[
\tilde{f}(w) \overset{\text{def}}{=} \int_0^\infty F(\alpha) d\alpha \quad (14.17)
\]

where the function \( F \) is defined as before.

- Note that it is not necessary in general to require \( w \in \mathbb{R}_+^E \) (i.e., we can take \( w \in \mathbb{R}^E \)) nor that \( f \) be non-negative, but it is a bit more involved. Above is the simple case.
Lovász extension, as integral

- Additional ways we can define the Lovász extension for any (not necessarily submodular) but normalized function $f$ include:

\begin{align*}
\tilde{f}(w) &= \sum_{i=1}^{m} w(e_i)f(e_{i} | E_{i-1}) \\
&= \sum_{i=1}^{m-1} f(E_i)(w(e_i) - w(e_{i+1})) + f(E)w(e_m) \\
&= \int_{\min\{w_1, \ldots, w_m\}}^{+\infty} f(\{w \geq \alpha\})d\alpha + f(E)\min\{w_1, \ldots, w_m\} \\
&= \int_{0}^{+\infty} f(\{w \geq \alpha\})d\alpha + \int_{-\infty}^{0} [f(\{w \geq \alpha\}) - f(E)]d\alpha
\end{align*}

Lovász extension properties

- Using the above, have the following (some of which we’ve seen):

**Theorem 14.2.3**

Let $f, g : 2^E \to \mathbb{R}$ be normalized ($f(\emptyset) = g(\emptyset) = 0$). Then

1. **Superposition of LE operator:** Given $f$ and $g$ with Lovász extensions $\tilde{f}$ and $\tilde{g}$ then $\tilde{f} + \tilde{g}$ is the Lovász extension of $f + g$ and $\lambda \tilde{f}$ is the Lovász extension of $\lambda f$ for $\lambda \in \mathbb{R}$.

2. If $w \in \mathbb{R}^E_+$ then $\tilde{f}(w) = \int_{0}^{+\infty} f(\{w \geq \alpha\})d\alpha$.

3. For $w \in \mathbb{R}^E$, and $\alpha \in \mathbb{R}$, we have $\tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E)$.

4. **Positive homogeneity:** i.e., $\tilde{f}(\alpha w) = \alpha \tilde{f}(w)$ for $\alpha \geq 0$.

5. For all $A \subseteq E$, $\tilde{f}(1_A) = f(A)$.

6. $f$ symmetric as in $f(A) = f(E \setminus A), \forall A$, then $\tilde{f}(w) = \tilde{f}(-w)$ ($\tilde{f}$ is even).

7. **Given partition $E^1 \cup E^2 \cup \cdots \cup E^k$ of $E$ and $w = \sum_{i=1}^{k} \gamma_i 1_{E^i}$ with $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k$, and with $E^{1:i} = E^1 \cup E^2 \cup \cdots \cup E^i$, then $\tilde{f}(w) = \sum_{i=1}^{k} \gamma_i f(E^1|E^{1:i-1}) = \sum_{i=1}^{k-1} f(E^{1:i})(\gamma_i - \gamma_{i+1}) + f(E)\gamma_k$.**
In fact, we have:

**Theorem 14.3.3**

Let $f$ be submodular and $\tilde{f}$ be its Lovász extension. Then
\[ \min \{ f(A) | A \subseteq E \} = \min_{w \in \{0,1\}^E} \tilde{f}(w) = \min_{w \in [0,1]^E} \tilde{f}(w). \]

**Proof.**

- First, since $\tilde{f}(1_A) = f(A)$, $\forall A \subseteq V$, we clearly have
  \[ \min \{ f(A) | A \subseteq V \} = \min_{w \in \{0,1\}^E} \tilde{f}(w) \geq \min_{w \in [0,1]^E} \tilde{f}(w). \]

- Next, consider any $w \in [0,1]^E$, sort elements $E = \{e_1, \ldots, e_m\}$ as $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m)$, define $E_i = \{e_1, \ldots, e_i\}$, and define $\lambda_m = w(e_m)$ and $\lambda_i = w(e_i) - w(e_{i+1})$ for $i \in \{1, \ldots, m-1\}$.

- Then, as we have seen, $w = \sum_i \lambda_i 1_{E_i}$ and $\lambda_i \geq 0$.

- Also, $\sum_i \lambda_i = w(e_1) \leq 1$. 

...
Minimizing $\tilde{f}$ vs. minimizing $f$

**...cont. proof of Thm. 14.3.3.**

- Note that since $f(\emptyset) = 0$, $\min \{ f(A) | A \subseteq E \} \leq 0$.
- Then we have
  \[
  \tilde{f}(w) = \int_0^1 f(\{w \geq \alpha\}) d\alpha = \sum_{i=1}^{m} \lambda_i f(E_i)
  \]
  \[
  \geq \sum_{i=1}^{m} \lambda_i \min_{A \subseteq E} f(A)
  \]
  \[
  \geq \min_{A \subseteq E} f(A)
  \]
- Thus, $\min \{ f(A) | A \subseteq E \} = \min_{w \in [0,1]^E} \tilde{f}(w)$.

Other minimizers based on min of $f$

- Let $w^* \in \arg\min \left\{ \tilde{f}(w) | w \in [0,1]^E \right\}$ and let $A^* \in \arg\min \{ f(A) | A \subseteq V \}$.
- Previous theorem states that $\tilde{f}(w^*) = f(A^*)$.
- Let $\lambda_i^*$ be the function weights and $E_i^*$ be the sets associated with $w^*$. From previous theorem, we have
  \[
  \tilde{f}(w^*) = \sum_i \lambda_i^* f(E_i^*) = f(A^*) = \min \{ f(A) | A \subseteq E \}
  \]
  and that $f(A^*) \leq f(E_i^*), \forall i$, and that $f(A^*) \leq 0$.
- Thus, since $w^* \in [0,1]^E$, each $0 \leq \lambda_i^* \leq 1$, we have for all $i$ such that $\lambda_i^* > 0$,
  \[
  f(E_i^*) = f(A^*)
  \]
  meaning such $E_i^*$ are also minimizers of $f$, and $\sum_i \lambda_i = 1$.
- Hence $w^*$ is in convex hull of incidence vectors of minimizers of $f$. 

Simple expressions for Lovász $E$ with $m = 2$, $E = \{1, 2\}$

- If $w_1 \geq w_2$, then
  \[
  \tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \\
  = (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\})
  \]  
  (14.1)

- If $w_1 \leq w_2$, then
  \[
  \tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \\
  = (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\})
  \]  
  (14.2)

- If $w_1 \geq w_2$, then
  \[
  \tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \\
  = (w_1 - w_2) f(\{1\}) + w_2 f(\{1, 2\})
  \]  
  (14.3)

- If $w_1 \leq w_2$, then
  \[
  \tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \\
  = (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\})
  \]  
  (14.4)

- A similar (symmetric) expression holds when $w_1 \leq w_2$. 

A similar (symmetric) expression holds when $w_1 \leq w_2$. 

\[
\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \\
= (w_2 - w_1) f(\{2\}) + w_1 f(\{1, 2\})
\]  
(14.5)

\[
\frac{1}{2} f(1)(w_1 - w_2) + \frac{1}{2} f(1)(w_1 - w_2)
\]  
(14.6)

\[
\frac{1}{2} f(\{1, 2\})(w_1 + w_2) - \frac{1}{2} f(\{1, 2\})(w_1 - w_2)
\]  
(14.7)

\[
\frac{1}{2} f(2)(w_1 - w_2) + \frac{1}{2} f(2)(w_2 - w_1)
\]  
(14.8)
This gives, for general \( w_1, w_2 \), that

\[
\tilde{f}(w) = \frac{1}{2} (f(\{1\}) + f(\{2\}) - f(\{1, 2\})) |w_1 - w_2| \quad (14.10)
\]

\[
+ \frac{1}{2} (f(\{1\}) - f(\{2\}) + f(\{1, 2\})) w_1 \quad (14.11)
\]

\[
+ \frac{1}{2} (-f(\{1\}) + f(\{2\}) + f(\{1, 2\})) w_2 \quad (14.12)
\]

\[
= -(f(\{1\}) + f(\{2\}) - f(\{1, 2\})) \min \{w_1, w_2\} \quad (14.13)
\]

\[
+ f(\{1\}) w_1 + f(\{2\}) w_2 \quad (14.14)
\]

Thus, if \( f(A) = H(X_A) \) is the entropy function, we have

\[
\tilde{f}(w) = H(e_1) w_1 + H(e_2) w_2 - I(e_1; e_2) \min \{w_1, w_2\} \text{ which must be convex in } w, \text{ where } I(e_1; e_2) \text{ is the mutual information.}
\]

This “simple” but general form of the Lovász extension with \( m = 2 \) can be useful.

Example: \( m = 2, E = \{1, 2\} \), contours

If \( w_1 \geq w_2 \), then

\[
\tilde{f}(w) = w_1 f(\{1\}) + w_2 f(\{2\}|\{1\}) \quad (14.15)
\]

- If \( w = (1, 0)/f(\{1\}) = (1/f(\{1\}), 0) \) then \( \tilde{f}(w) = 1 \).
- If \( w = (1, 1)/f(\{1, 2\}) \) then \( \tilde{f}(w) = 1 \).

If \( w_1 \leq w_2 \), then

\[
\tilde{f}(w) = w_2 f(\{2\}) + w_1 f(\{1\}|\{2\}) \quad (14.16)
\]

- If \( w = (0, 1)/f(\{2\}) = (0, 1/f(\{2\})) \) then \( \tilde{f}(w) = 1 \).
- If \( w = (1, 1)/f(\{1, 2\}) \) then \( \tilde{f}(w) = 1 \).

With this we can plot contours of the form \( \{w \in \mathbb{R}^2 : \tilde{f}(w) = 1\} \) with marked points of the form \( 1_A \times 1/f(A) \).
**Example:** $m = 2, E = \{1, 2\}$

- Contour plot of $m = 2$ Lovász extension (from Bach-2011).

```
(0, 1)/f(\{2\})
w_2 > w_1
(1, 1)/f(\{1, 2\})

\{w : f(\tilde{w}) = 1\}
```

- $w_2 > w_1$  
- $(0, 1)/f(\{2\})$  
- $(1, 1)/f(\{1, 2\})$

- $w_1 > w_2$

- Q: does $f$ appear to be a polymatroid function?

---

**Example: **$m = 3, E = \{1, 2, 3\}$

- In order to visualize in 3D, we make a few simplifications.
- Consider any submodular $f'$ and $x \in B_{f'}$. Then
  \[
  f(A) = f'(A) - x(A) \text{ is submodular, and moreover } f(E) = f'(E) - x(E) = 0.
  \]
- Hence, from $\tilde{f}(w + \alpha 1_E) = \tilde{f}(w) + \alpha f(E)$, we have that
  \[
  \tilde{f}(w + \alpha 1_E) = \tilde{f}(w).
  \]
- Thus, we can look “down” on the contour plot of the Lovász extension, \{w : \tilde{f}(w) = 1\}, from a vantage point right on the line \{x : x = \alpha 1_E, \alpha > 0\} since moving in direction $1_E$ changes nothing.
Example: \( m = 3, \ E = \{1, 2, 3\} \)

- Example 1 (from Bach-2011): \( f(A) = 1_{|A| \in \{1, 2\}} = \min \{|A|, 1\} + \min \{|E \setminus A|, 1\} - 1 \) is submodular, and \( \tilde{f}(w) = \max_{k \in \{1, 2, 3\}} w_k - \min_{k \in \{1, 2, 3\}} w_k \).

- Example 2 (from Bach-2011):
  \[
  f(A) = |1_{1 \in A} - 1_{2 \in A}| + |1_{2 \in A} - 1_{3 \in A}|
  \]
  This gives a “total variation” function for the Lovász extension, with \( \tilde{f}(w) = |w_1 - w_2| + |w_2 - w_3| \), a prior to prefer piecewise-constant signals.
Total Variation Example

From “Nonlinear total variation based noise removal algorithms” Rudin, Osher, and Fatemi, 1992. Top left original, bottom right total variation.

Example: Lovász extension of concave over modular

- Let \( m : E \to \mathbb{R}_+ \) be a modular function and define 
  \[ f(A) = g(m(A)) \] 
  where \( g \) is concave. Then \( f \) is submodular.
- Let \( M_j = \sum_{i=1}^j m(e_i) \)
- \( \tilde{f}(w) \) is given as
  \[ \tilde{f}(w) = \sum_{i=1}^m w(e_i)(g(M_i) - g(M_{i-1})) \]  \hspace{1cm} (14.17)
- And if \( m(A) = |A| \), we get
  \[ \tilde{f}(w) = \sum_{i=1}^m w(e_i)(g(i) - g(i-1)) \]  \hspace{1cm} (14.18)
Example: Lovász extension and cut functions

- Cut Function: Given a non-negative weighted graph $G = (V, E, m)$ where $m : E \to \mathbb{R}_+$ is a modular function over the edges, we know from Lecture 2 that $f : 2^V \to \mathbb{R}_+$ with $f(X) = m(\Gamma(X))$ where $\Gamma(X) = \{(u, v) | (u, v) \in E, u \in X, v \in V \setminus X\}$ is non-monotone submodular.

- Simple way to write it, with $m_{ij} = m((i, j))$:

$$f(X) = \sum_{i \in X, j \in V \setminus X} m_{ij} \tag{14.19}$$

- Exercise: show that Lovász extension of graph cut may be written as:

$$\tilde{f}(w) = \sum_{i, j \in V} m_{ij} \max \{ (w_i - w_j), 0 \} \tag{14.20}$$

where elements are ordered as usual, $w_1 \geq w_2 \geq \cdots \geq w_n$.

- This is also a form of “total variation”

A few more Lovász extension examples

Some additional submodular functions and their Lovász extensions, where $w(e_1) \geq w(e_2) \geq \cdots \geq w(e_m) \geq 0$. Let $W_k \triangleq \sum_{i=1}^k w(e_i)$.

<table>
<thead>
<tr>
<th>$f(A)$</th>
<th>$\tilde{f}(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>A</td>
</tr>
<tr>
<td>$\min(</td>
<td>A</td>
</tr>
<tr>
<td>$\min(</td>
<td>A</td>
</tr>
<tr>
<td>$\min(</td>
<td>A</td>
</tr>
<tr>
<td>$\min(</td>
<td>A</td>
</tr>
<tr>
<td>$\min(</td>
<td>A</td>
</tr>
</tbody>
</table>

(thanks to K. Narayanan).
Lovász extension and norms

- In general, Lovász extension can be useful to define various norms of the form \( \|w\|_{\tilde{f}} = \tilde{f}(|w|) \), which renders the function symmetric about all orthants (i.e., \( \|w\|_{\tilde{f}} = \|b \circ w\|_{\tilde{f}} \) where \( b \in \{-1, 1\}^m \) and \( \circ \) is element-wise multiplication).

- Simple example. The Lovász extension of the modular function \( f(A) = |A| \) is the \( \ell_1 \) norm, and the Lovász extension of the modular function \( f(A) = m(A) \) is the weighted \( \ell_1 \) norm.

- With more general submodular functions, one can generate a large and interesting variety of norms, all of which have polyhedral contours (unlike, say, something like the \( \ell_2 \) norm).

- Hence, not all norms come from the Lovász extension of some submodular function.

- Similarly, not all convex functions are the Lovász extension of some submodular function.

- Bach-2011 has a complete discussion of this.

Summary important definitions so far: tight, dep, & sat

- \( x \)-tight sets: For \( x \in P_f \), \( D(x) = \{A \subseteq E : x(A) = f(A)\} \).

- Polymatroid closure/maximal \( x \)-tight set: For \( x \in P_f \),
  \[ \text{sat}(x) = \bigcup\{A : A \in D(x)\} = \{e : e \in E, \forall \alpha > 0, x + \alpha e \notin P_f\} \]

- Saturation capacity: for \( x \in P_f \), \( 0 \leq \hat{c}(x; e) = \min \{f(A) - x(A) | \forall A \ni e\} = \max \{\alpha : \alpha \in \mathbb{R}, x + \alpha e \in P_f\} \)

- Recall: \( \text{sat}(x) = \{e : \hat{c}(x; e) = 0\} \) and \( E \setminus \text{sat}(x) = \{e : \hat{c}(x; e) > 0\} \).

- \( e \)-containing \( x \)-tight sets: For \( x \in P_f \),
  \[ D(x, e) = \{A : e \in A \subseteq E, x(A) = f(A)\} \subseteq D(x) \]

- Minimal \( e \)-containing \( x \)-tight set/polymatroidal fundamental circuit/: For \( x \in P_f \),
  \[ \text{dep}(x, e) = \begin{cases} \bigcap \{A : e \in A \subseteq E, x(A) = f(A)\} & \text{if } e \in \text{sat}(x) \\ \emptyset & \text{else} \end{cases} = \{e' : \exists \alpha > 0, \text{ s.t. } x + \alpha (1_e - 1_{e'}) \in P_f\} \]
dep and sat in a lattice

- The picture on the right summarizes the relationships between the lattices and sublattices.
- Note, \( \bigcap_e \text{dep}(x, e) = \text{dep}(x) \).

Summary of \( \text{sat}, \text{dep} \)

- For \( x \in P_f \), \( \text{sat}(x) \) (span, closure) is the maximal saturated \((x\text{-tight})\) set w.r.t. \( x \). I.e., \( \text{sat}(x) = \{ e : e \in E, \forall \alpha > 0, x + \alpha 1_e \notin P_f \} \). That is,
  \[
  \text{cl}(x) \overset{\text{def}}{=} \text{sat}(x) \triangleq \bigcup \{ A : A \in \mathcal{D}(x) \} = \bigcup \{ A : A \subseteq E, x(A) = f(A) \} = \{ e : e \in E, \forall \alpha > 0, x + \alpha 1_e \notin P_f \}
  \]

- For \( e \in \text{sat}(x) \), \( \text{dep}(x, e) \) (fundamental circuit) is the minimal \((\text{common})\) saturated \((x\text{-tight})\) set w.r.t. \( x \) containing \( e \). That is,
  \[
  \text{dep}(x, e) = \begin{cases} 
  \bigcap \{ A : e \in A \subseteq E, x(A) = f(A) \} & \text{if } e \in \text{sat}(x) \\
  \emptyset & \text{else} \\
  \{ e' : \exists \alpha > 0, \text{ s.t. } x + \alpha (1_e - 1_{e'}) \in P_f \} & \end{cases}
  \]
Consider \( x \in P_f \), and consider the following set
\[
\text{DEP}(x) = \{ \text{dep}(x, e) : e \in \text{sat}(x) \} \tag{14.21}
\]
So \( \text{DEP}(x) \) is a set of sets, each element of \( \text{DEP}(x) \) is the \( \text{dep}(x, e) \) valuation for some \( e \in \text{sat}(x) \).

Moreover, define a partial order on \( \text{DEP}(x) \) as follows: if \( A, B \in \text{DEP}(x) \), then \( A \preceq B \) iff \( A \subseteq B \).

We’re going to use this partial order to define a partial order on all elements of \( \text{sat}(x) \).

Now recall \( D(x) = \{ A : x(A) = f(A) \} \) forms a distributive lattice. What is the natural partial order?

Now in any distributive lattice \( L \), consider its join-irreducibles \( \mathcal{J} \) (i.e., any element \( A \in \mathcal{J} \) can’t be represented as a join of any other two elements in \( L \)).

Fact (see Birkhoff, 1969) if the lattice has length \( n \), then \( \mathcal{J} \) will have exactly \( n \) elements (in the Boolean case, these are atoms/ground elements), and each element in \( \mathcal{J} \) is partially ordered by the lattice partial order.

Moreover, any element can be “generated” by joining the join-irreducible elements.
dep and partial order

- Now any element in \( \text{DEP}(x) \) (for \( x \) extreme) can’t be represented by the join of two other elements in \( \text{DEP}(x) \).
- Specifically, let \( e, a, b \in E \) be such that \( \text{dep}(x, e), \text{dep}(x, a), \text{dep}(x, b) \in \text{DEP}(x) \). Then we can’t have \( \text{dep}(x, e) = \text{dep}(x, a) \cup \text{dep}(x, b) \), unless either \( \text{dep}(x, e) = \text{dep}(x, b) \) or \( \text{dep}(x, e) = \text{dep}(x, a) \), meaning they are join irreducibles.
- The reason is since the minimal tight sets containing \( e \) would not be generated by merging two minimal tight sets containing, say, \( a \), and \( b \), where all of \( a, b, e \) are unequal.
- Thus, considering \( D(x) \) as a distributed lattice, then \( \text{DEP}(x) \) are the join-irreducibles.
- And the order \( \preceq \) defined earlier is the natural order w.r.t. this lattice and its join-irreducibles.

Let \( x \in P_f \) again be an extreme point, and let it be generated by an ordering of \( B = (e_1, e_2, \ldots, e_k) \subseteq E \) with \( B_i = (b_i, b_2, \ldots, b_i) \), \( i \leq k \) a prefix order w.r.t. ordered items \( B \) (\( B \) and \( B_i, \forall i \) are ordered sets).
- Recall, the equation for \( x \) is of the form \( x(e) = 0 \) for some \( e \) and \( x(A) = f(A) \) for some \( A \) (see earlier). Specifically, we have that \( x(E \setminus B) = 0 \) and, for \( i = 1 \ldots k \), \( x(B_i) = f(B_i) \).
- Thus, each of \( B_i \) is a tight set.
- We also have that \( \text{supp}(x) \subseteq B \).
dep and partial order

- Thus, for any \( d, e \in \text{supp}(x) \subseteq B \), there is a tight set containing one but not the other. Specifically, let \( d = e_i \) and \( e = e_j \) with \( i < j \). Then non-zero \( B_i \) (i.e., \( B_i \cap \text{supp}(x) \)) contains \( d \) but not \( e \).
- So there is a tight set (namely \( B_i \)) that contains \( d = e_i \) but not \( e = e_j \) with \( j > i \) (note, vice versa is not true).
- Thus, for any \( d, e \in \text{supp}(x) \subseteq B \), we have \( \text{dep}(x, d) \neq \text{dep}(x, e) \).
- That is, \( B_i \) is a tight set with \( d \) but not with \( e \). Since \( \text{dep}(x, d) \subseteq B_i \), we thus have \( e \notin \text{dep}(x, d) \), but of course \( e \in \text{dep}(x, e) \), so \( \text{dep}(x, d) \neq \text{dep}(x, e) \).
- Moreover, for any \( d \in B \), it might be that \( \text{dep}(x, d) = B_i \) where \( d = e_i \). This point is further clarified in the next slide.

dep and partial order (slight digression)

- I.e., \( x \) is extreme generated by \( B \Rightarrow B_i \) is a tight set containing \( e_i \).
- For any \( j < i \), \( B_j \) does not contain \( e_i \).
- Thus, \( \text{dep}(x, e_i) \) (minimal tight \( e_i \)-containing set) might equal \( B_i \).
- On the other hand, consider the extreme vector \( x^{(i)} \in \mathbb{R}^E \) with
  \[
x^{(i)}(e) = \begin{cases} x(e) & \text{if } e \in B_i \\ 0 & \text{else} \end{cases}
\]
  so \( x^{(i)} \) is just the extreme vector generated by the ordered set \( B_i \).
- Therefore, \( B_j \) for \( j \leq i \) are tight w.r.t. \( x^{(i)} \).
- Could be another ordered set (say \( B^{(i)} \), which is \( B_i \) permuted) that also generates \( x^{(i)} \). Let \( B_j^{(i)} \), \( j \leq i \) be the first \( j \) elements in \( B^{(i)} \).
- In \( B^{(i)} \), \( e_i \) might come at a position \( j < i \), so \( B_j^{(i)} \) is tight and containing \( e_i \), and \( \text{dep}(x, e_i) \) might equal \( B_j^{(i)} \), with \( B_j^{(i)} \subset B_i \).
Hence, $B_i$ is an $x$-tight set with $e_i$ at position $|B_i| = i$, and $B_j^{(i)}$ is a permutation of $B_i$ with $e_i$ at position $j < i$, and is also an $x$-tight set.

On the other hand, $B_j^{(i)} \not\subseteq \text{dep}(x, e_i)$ and $B_i \not\subseteq \text{dep}(x, e_i)$ due to $\text{dep}(x, e_i)$’s minimality.

Therefore, in general, $\text{dep}(x, e_i) \subseteq B_i$ and $\text{dep}(x, e_i) \subseteq B_j^{(i)}$.

And this is true regardless of the permutation of $B_i$, as long as it generates $x^{(i)}$.

Now, while $\text{dep}(x, e_i) \subseteq B_i$, we can be a bit more explicit.

Let $B(x)$ be set of permutations of $B$ that generate $x$.

For $e \in B$ and $B' \in B(x)$, let $1 \leq e(B') \leq |B'|$ be $e$’s position in $B'$.

Then $\text{dep}(x, e_i) = B_j^{e_i}$ where

$$B_j^{e_i} \in \arg\min_{B' \in B(x)} e_i(B')$$

(14.23)

is an ordered set, and $j$ is the position of $e_i$ in $B_j^{e_i}$, i.e., $j = e_i(B_j^{e_i})$.

This follows from iff relationship between extremal points and greedy algorithm, and since $\text{dep}(x, e_i)$ is the unique “0” element of a distributive lattice.

Then since $\text{dep}(x, e_i) = B_j^{e_i}$, and it is the unique minimal $e_i$-containing $x$-tight set, we also have

$$|\arg\min_{B' \in B(x)} e_i(B')| = 1,$$

meaning

$$\{B_j^{e_i}\} = \arg\min_{B' \in B(x)} e_i(B')$$

(14.24)
B is an ordered set of size $k \leq m = |E|$, and $B_j$ is an ordered set consisting of $j$-element prefix of $B$, so $|B_j| = j$.

For ordered set $B$, and $e \in B$, we have $e(B)$ is the index of $e$ within $B$. I.e., if $B = (e_1, e_2, \ldots, e_k)$ then $e_i(B) = i$ for $1 \leq i \leq k$.

$B$ generates $x \in \mathbb{R}^E$, meaning $x(e_i) = f(e_i|B_{i-1})$ for $e_i \in B$.

$B(x)$ is the set of permutations of $B$ that generate $x$, meaning for any $B' \in B(x)$, we have $x(e_i) = f(e_i|B'_{i-1})$ for $e_i \in B$.

$B^e$ is the permutation within $B(x)$ where $e$ occurs earliest. I.e., $e(B^e) \leq e(B')$ for any $B' \in B(x)$.

$B^e_j$ is the $j$-element prefix of $B^e$.

$B^e$ is just $B^e_i$ for $e = e_i$. Same with $B^e_j$ and $B^e_i$.

Previous slide, we argued that for $e \in \text{sat}(x)$, $\text{dep}(x, e) = B^e_j$ with $j = e(B^e)$ (could also write this as $\text{dep}(x, e) = B^e_{e(B^e)}$).

Dep and partial order (slight digression)

Now, for $d, e \in \text{sat}(x)$, we have that

$$\text{dep}(x, d) = B^d_i \subset \text{dep}(x, e) = B^e_j \iff d \in \text{dep}(x, e) \quad (14.25)$$

where $i = d(B^d)$ and $j = e(B^e)$.

Proof:

Clearly, $\text{dep}(x, d) \subset \text{dep}(x, e) \Rightarrow d \in \text{dep}(x, e)$.

Conversely, $d \in \text{dep}(x, e) = B^e_j$ means $\text{dep}(x, d) \subset B^e_k$ where $k = d(B^e)$ is the position of $d$ in $B^e$ (since $B^e_k$ is a tight set containing $d$).

It must be that $k < j$ (since $B^e_j$ is the smallest tight set containing $e$, and the $j$’th position of $B^e_j$ is $e$ whose removal doesn’t remove $d$, leaving another tight set $B^e_j - e$ containing $d$).

Therefore, $\text{dep}(x, d) \subseteq B^e_k \subseteq B^e_j - e \subset B^e_j = \text{dep}(x, e)$.

As a consequence, if $d \in \text{dep}(x, e)$ then $e \notin \text{dep}(x, d)$. 
The next three slides are review from lecture 11.

Tightness of \( \text{supp} \) at polymatroidal extreme point

- Now, \( \text{sat}(x) \) is tight, and corresponds to the largest member of the distributive lattice \( \mathcal{D}(x) = \{ A : x(A) = f(A) \} \) of tight sets.
- \( \text{supp}(x) \) is not necessarily tight for an arbitrary \( x \).
- When \( x \) is an extremal point, however, \( \text{supp}(x) \) is tight, meaning \( x(\text{supp}(x)) = f(\text{supp}(x)) \). Why?
  1. Extremal points are defined as a system of equalities of the form \( x(E_i) = f(E_i) \) for \( 1 \leq i \leq k \leq |E| \), for some \( k \), as we saw earlier in class. Hence, any \( e_i \in \text{supp}(x) \) has \( x(e_i) = f(e_i | E_i - 1) > 0 \).
  2. Now, for \( 1 \leq i \leq k \), if \( e_i \notin \text{supp}(x) \), \( x(E_k) = x(E_k - e_i) \)
  3. Also, for \( 1 \leq i \leq k \), if \( e_i \notin \text{supp}(x) \), then \( 0 = f(e_i | E_i - 1) \geq f(e_i | E_k - e_i) = f(E_k | E_k - e_i) \geq 0 \) since monotone submodular, hence we have \( f(E_k) = f(E_k - e_i) \).
  4. We can keep removing elements \( e_i \notin \text{supp}(x) \) and we’re left with \( f(E_k \cap \text{supp}(x)) = x(E_k \cap \text{supp}(x)) \) for any \( k \).
  5. Hence \( \text{supp}(x) \) is tight when \( x \) is extremal.
- Since \( \text{supp}(x) \) is tight, we immediately have that \( \text{sat}(x) \supseteq \text{supp}(x) \).
more Lovász extension

Lovász extension examples

Partial order of extreme points

Summary

Scratch

**supp vs. sat equality**

- For \( x \in P_f \), with \( x \) extremal, is \( \text{supp}(x) = \text{sat}(x) \)?
- Consider an example case where disjoint \( X, Y \subseteq E \), we have \( f(X) = f(Y) = f(X \cup Y) \) (meaning “perfect dependence” or full redundancy, so gains are not strictly positive).
- Suppose \( x \in P_f \) has \( x(X) > 0 \) but \( x(V \setminus X) = 0 \) and so \( x(Y) = 0 \).
- Then \( \text{supp}(x) = X \)
- \( \text{sat}(x) = \bigcup\{ A : x(A) = f(A) \} \) and since \( x(X \cup Y) = x(X) = f(X) = f(X \cup Y) \), here, \( \text{sat}(x) \supseteq X \cup Y \).
- In general, for extremal \( x \), \( \text{sat}(x) \supseteq \text{supp}(x) \) (see later).
- Also, recall \( \text{sat}(x) \) is like span/closure but \( \text{supp}(x) \) is more like indication. So this is similar to \( \text{span}(A) \supseteq A \).
- For modular functions, they are always equal (e.g., think of “hyperrectangular” polymatroids).

**supp, sat, extremal \( x \), perfect dependence**

- In general, for extremal \( x \), \( \text{sat}(x) \supseteq \text{supp}(x) \).
- Now, for any \( e \in E \setminus \text{supp}(x) \), we clearly have \( x(\text{supp}(x) + e) = x(\text{supp}(x)) \) since \( x(e) = 0 \).
- On the other hand, for \( e_i \in \text{sat}(x) \setminus \text{supp}(x) \), we have perfect dependence, i.e., \( f(\text{supp}(x) + e_i) = f(\text{supp}(x)) \). Indeed:
  - \( \text{sat}(x) \) is tight, as is \( \text{supp}(x) \), and hence \( f(\text{sat}(x)) = x(\text{sat}(x)) = x(\text{supp}(x)) = f(\text{supp}(x)) \).
  - Therefore, \( f(\text{sat}(x) | \text{supp}(x)) = 0 \).
  - But by the above, and monotonicity, we have for \( e \in \text{sat}(x) \setminus \text{supp}(x) \), that \( 0 = f(\text{sat}(x) | \text{supp}(x)) \geq f(e | \text{supp}(x)) \geq 0 \).
  - Hence \( f(e | \text{supp}(x)) = 0 \), and moreover \( f(e + \text{supp}(x)) = x(e + \text{supp}(x)) = x(\text{supp}(x)) = f(\text{supp}(x)) \).
- Thus, for any extremal \( x \), with \( \text{sat}(x) \supseteq \text{supp}(x) \), we see that for \( e \in \text{sat}(x) \setminus \text{supp}(x) \), we have \( \text{supp}(x) + e \) is also tight.
- Note also, for any \( A \subseteq \text{sat}(x) \setminus \text{supp}(x) \), we have \( f(A | \text{supp}(x)) = 0 \).
Recall, for polymatroidal $f$, we saw earlier that for each $e \in \text{sat}(x) \setminus \text{supp}(x)$, the set $\text{supp}(x) + e$ is also tight.

Now for any point $a, b \in \text{sat}(x) \setminus \text{supp}(x)$, we have that $\text{dep}(x, a) \neq \text{dep}(x, b)$

This follows, since the minimal tight set containing $a$ would never contain $b$ since $f(\text{supp}(x) + a + b) = x(\text{supp}(x) + a + b) = x(\text{supp}(x) + a) = f(\text{supp}(x) + a)$ (and in this case, vice versa).

I.e., in such case, we can have for $a \in \text{sat}(x) \setminus \text{supp}(x)$, $\text{dep}(x, a) = B'_j + a$ for some ordering $B' \in \mathcal{B}(x)$, and for some $j$, the smallest $j$ such that $f(B'_j + a) = f(B'_j)$, and note that $a \notin B_j$.

This gives further support to the phrase “dependence function”, namely $\text{dep}(x, a) \setminus \{a\} = B'_j$ is the smallest set that renders $a$ dependent (again, like the fundamental circuit of a matroid).

So, for $a, b \in \text{supp}(x)$, $\text{dep}(x, a) \neq \text{dep}(x, b)$ (from Slide 32).

And for $a, b \in \text{sat}(x) \setminus \text{supp}(x)$, $\text{dep}(x, a) \neq \text{dep}(x, b)$ (from Slides 42)

What about $a \in \text{sat}(x) \setminus \text{supp}(x)$ and $b \in \text{supp}(x)$? The minimal tight set containing $b$ would never contain $a$ since $f(\text{supp}(x)) = x(\text{supp}(x))$ and $f(\text{supp}(x) + a) = x(\text{supp}(x) + a)$, so $\text{dep}(x, b)$ has no need for $a$.

Hence, for any $a, b \in \text{sat}(x)$, we have $\text{dep}(x, a) \neq \text{dep}(x, b)$.

Thus, we have 1-1 mapping between all elements of $\text{sat}(x)$ and $\text{DEP}(x) = \{\text{dep}(x, e) : e \in \text{sat}(x)\}$.
dep and partial order

- Therefore, the partial order on \( \text{DEP}(x) \) can be used to define a partial order on \( \text{sat}(x) \).
- Now, for \( d, e \in \text{sat}(x) \), when can we have that \( \text{dep}(x, d) \subseteq \text{dep}(x, e) \)?
- We already saw this on Slide 35, where this happens iff \( d \in \text{dep}(x, e) \), for \( d \neq e \).
- Recall also from Slide 35, if \( d \in \text{dep}(x, e) \) then \( e \notin \text{dep}(x, d) \).
- Thus, we can define a partial order on the elements of \( \text{sat}(x) \) as follows:

\[ \text{Definition 14.5.1 (partial order on elements of } \text{sat}(x) \text{)} \]

For \( d, e \in \text{sat}(x) \), we have

\[ d \preceq e \iff d \in \text{dep}(x, e) \] (14.26)

- Thus, we have just proven

\[ \text{Theorem 14.5.2} \]

If \( x \in P_f \) is an extreme point, then \( \preceq \) is a partial order on \( \text{sat}(x) \) where for \( a, e \in \text{sat}(x) \), the order \( \preceq \) is defined by: \( a \preceq e \) iff \( a \in \text{dep}(x, e) \).

- In other words, \( x \in P_f \) is an extreme point \( \Rightarrow \) the construct

  \[ [\text{for } a, e \in \text{sat}(x), a \preceq e \text{ iff } a \in \text{dep}(x, e)] \]

defines a partial order.

- We can show a stronger result that extreme points are characterized by this construct. I.e., following converse can be shown:

\[ \text{Theorem 14.5.3} \]

\( x \in P_f \) is an extreme point iff \( \text{supp}(x) \subseteq \text{sat}(x) \) and \( \text{dep}(x, a) \neq \text{dep}(x, b) \) for every pair of distinct points \( a, b \in \text{sat}(x) \) (meaning we can define a partial order on \( \text{sat}(x) \) as above).
dep, strict submodularity, and total order

- If \( f(A) + f(B) > f(A \cup B) + f(A \cap B), \forall A, B : A \not\subseteq B, B \not\subseteq A \) (i.e., \( f \) is strictly submodular) then the above order \( \preceq \) is a total order on \( \text{sat}(x) \).

- Strictly submodular is same condition as \( f(e|A) > f(e|B) \) for all \( A \subset B \subseteq E \setminus \{e\} \).

- Now our goal is to be able to, given an extreme point \( x \in P_f \) characterize \( \preceq \), and in particular generate \( \preceq \) and thus characterize all orderings that generate \( x \).

**Definition 14.5.4**

Given a partial order \( \preceq \), and an ordered set \( B = (e_1, e_2, \ldots, e_k) \), then \( B \) is compatible with \( \preceq \) if \( i < j \) whenever \( e_i \preceq e_j \) (\( \equiv e_i \in \text{dep}(x, e_j) \)) and where \( e_i, e_j \) are distinct.

That is, \( B \) is compatible with \( \preceq \) if, given distinct \( e_i, e_j \),

\[ e_i \preceq e_j \Rightarrow i < j. \]

**Theorem 14.5.5**

Let \( x \) be an extreme point of \( P_f \) and \( \preceq \) be its partial order. Let \( B \subseteq E \) be an ordered set. Then \( B \) generates \( x \) using the greedy algorithm iff we have \( \text{supp}(x) \subseteq B \subseteq \text{sat}(x) \) and \( B \) is compatible with \( \preceq \).

**Proof.**

- Generate \( \Rightarrow \) Compatible: Let \( B = \{b_1, \ldots, b_k\} \) generate \( x \)
  - Then \( \text{supp}(x) \subseteq B \).
  - Also, since \( B \) is tight, \( B \in \mathcal{D}(x) \), so \( B \subseteq \text{sat}(x) \).
  - Moreover, \( B_j \in \mathcal{D}(x) \) (for \( 1 \leq j \leq |B| \)), so that \( \text{dep}(x, e_j) \subseteq B_j \) for \( e_j \) the \( j^{th} \) element of \( B \). Hence, any \( e_i \in \text{dep}(x, e_j) \subseteq B_j \) can’t have \( j = i \) and must have \( i < j \). So any \( e_i \preceq e_j \) has \( i < j \).
  - But \( g = e_i \not\in B_j \) means \( g \not\in \text{dep}(x, e_j) \), or \( g \not\preceq e_j \), and this also requires \( i > j \).
  - Hence, \( B \) is compatible with \( \preceq \).
the partial order of extreme points

**Theorem 14.5.5**

Let $x$ be an extreme point of $P_f$ and $\preceq$ be its partial order. Let $B \subseteq E$ be an ordered set. Then $B$ generates $x$ using the greedy algorithm iff we have $\text{supp}(x) \subseteq B \subseteq \text{sat}(x)$ and $B$ is compatible with $\preceq$.

**Proof.**

- Conversely (Compatible $\Rightarrow$ Generate): Suppose ordering $B$ is compatible with $\preceq$ and that $\text{supp}(x) \subseteq B \subseteq \text{sat}(x)$.

- For each $j$ (with $1 \leq j \leq \vert B \vert$), consider $e_\ell$ at position $\ell$ in $B_j$, and consider any $e \in \text{dep}(x, e_\ell)$ so $e \preceq e_\ell$. Compatibility means we must have $i < \ell$ where $i$ is the position within $B_j$ of $e$, so $e \in B_j$.

- Hence, for each $e \in B_j$, we have $\text{dep}(x, e) \subseteq B_j$.

- Thus, $B_j$ is the union of tight sets (each of $\text{dep}(x, e)$ is tight), so that $B_j$ is also tight (unions of tight sets are tight).

- That is, we have $x(B_j) = f(B_j), 1 \leq j \leq \vert B \vert$.

- Thus $x$ is generated by greedy using ordering given in $B$.

**Corollary 14.5.6**

If $x$ is an extreme point of $P_f$ and $B \subseteq E$ is given such that $\text{supp}(x) \subseteq B \subseteq \text{sat}(x)$, then $x$ is generated using greedy by some ordering of $B$.

- this is a more satisfying way to, given an extreme point, show that the greedy algorithm can generate it than to resort to the polyhedral $cv = \max(cx : x \in P_f)$ for an appropriate direction $c$.

- Moreover, we can produce an efficient $O(|E|^2)$ algorithm that can produce $\preceq$ for any extreme point $x$ of $P_f$.

- The algorithm does so by, for each $e \in \text{sat}(x)$, producing the sets $\text{dep}(x, e)$ (or otherwise terminating by saying that $x$ is not an extreme point).

- Thus, extreme point testing is fundamentally computationally simpler than arbitrary membership testing (recall, to test if $x \in P_f$ in general, we needed submodular function minimization).
Extreme point testing and partial order generation

**input**: Vector $x \in \mathbb{R}^E$, polymatroid function $f$ on $E$.

**output**: That $x$ is not extreme point, or if it is, minimal tight sets $\text{dep}(x, e)$ for $e \in \text{sat}(x)$ thus defining $\preceq$. Moreover, $\text{dep}(x, e_j) = A_j$ for $1 \leq j \leq n$ where $n = |\text{sat}(x)|$.

```
1  j ← 0 ; B ← ∅ ;
2  while true do
3      j ← j + 1 ;
4      if $\exists e \in E \setminus B$ with $x(B + e) = f(B + e)$ then
5          B ← B + e ; e_j ← e ;
6      else
7          STOP, if supp($x$) $\subseteq B$ then $x$ is extreme, otherwise not.
8      A_j ← B ; k ← j − 1 ;
9      while $x(A_j - e_k) = f(A_j - e_k)$ and $k > 0$ do
10         A_j ← A_j − e_k ; k ← k − 1 ;
```

**On partial order algorithm**

- Lines 4-5 just arbitrarily adds elements, maintaining tightness of $B$.
- Lines 9-10 remove elements from $A_j$ while retaining tightness (thus achieving $\text{dep}(x, e_j)$).
- Algorithm uses $f$ only to test tightness of a set relative to a vector $x$, nothing more (i.e., line 4 could be a query on if $B + e$ is tight).
- Line 7 reports non-extreme if condition supp($x$) $\subseteq B \subseteq \text{sat}(x)$ violated ($\text{sat}(x)$ condition implicit).
- Line 1 is $O(|E|)$, and nested lines 4 (and 9) are each $O(|E|)$, so algorithm runs in $O(|E|^2)$ doing that many function evaluations.
- Thus, extreme point testing is fundamentally computationally simpler than arbitrary membership testing (recall, to test if $x \in P_f$ in general, we needed submodular function minimization).
- To determine, only, if a given $x$ is extreme, we can delete lines 8-10 (having same run time).
- If desired, we can generate all orderings consistent with a partial ordering using an algorithm by Knuth/Szwarcfiter-1974.
- To prove correctness, we need a few theorems.
Maximal in a tight set

**Theorem 14.5.7**

*Given an extreme point \( x \in P_f \), with \( A \) tight for \( x \), and if given order \( \preceq \) element \( e \in A \) is maximal, then \( A - e \) is also tight.*

**Proof.**

- Since \( A \) is tight, \( \forall a \in A, \text{dep}(x, a) \subseteq A \), and \( \bigcup_{a \in A} \text{dep}(x, a) = A \).
- If \( e \) is maximal in \( A \) w.r.t. \( \preceq \), then there exists no \( a \in A \setminus \{e\} \), such that \( e \preceq a \) (i.e., \( e \in \text{dep}(x, a) \)). I.e., \( e \notin \text{dep}(x, a) \) for any \( a \in A \setminus \{e\} \).
- Thus, \( \text{dep}(x, a) \subseteq A \setminus \{e\} \) for all \( a \in A \setminus \{e\} \).
- Now, since \( \text{dep}(x, a) \) is the smallest \( x \)-tight set containing \( a \) and \( \text{dep}(x, a) \subseteq A \setminus \{e\} \), we have
  \[
  \bigcup_{a \in A\setminus\{e\}} \text{dep}(x, a) = A \setminus \{e\} 
  \]  
  (14.27)
- Hence, \( A \setminus \{e\} \) is therefore also tight.

We also have

**Corollary 14.5.8**

*For all \( e \in \text{sat}(x) \), we have that \( \text{dep}(x, e) \setminus \{e\} \) is also tight.*

**Proof.**

- \( \text{dep}(x, e) \) is tight and \( e \) is maximal within \( \text{dep}(x, e) \).
- This theorem and corollary allow us to prove that the above algorithm gives us not only the minimum sets containing \( e \) but the minimum tight sets with \( e \), i.e., \( \text{dep}(x, e) \).
The output of Algorithm 1 is as asserted in the statement of the algorithm.

Proof.

- First, we prove that $x$ is an extreme point iff the algorithm terminates with $\text{supp}(x) \subseteq B$.
  - The algorithm maintains tightness for all sets $B$ so constructed.
  - If we terminate with $\text{supp}(x) \subseteq B$, then the resulting ordering $B = \{e_1, e_2, \ldots, e_k\}$ generates all of $x$ due to the tight equations $x(E_i) = f(E_i)$, $1 \leq i \leq k$ so $x$ is extreme.
  - Conversely, suppose $\text{supp}(x) \setminus B \neq \emptyset$ but $x$ is still extreme.
  - Hence, for any $a \in \text{supp}(x) \setminus B$, the set $B + a$ is not tight.
  - Now, $\text{supp}(x) \setminus B$ has a minimal element according to $\preceq$, say $b$.
  - So, $\exists c \in \text{supp}(x) \setminus B$ with $c \in \text{dep}(x, b)$ (i.e., $c \preceq b$). Thus $\text{dep}(x, b) \cap (\text{sat}(x) \setminus B) = \emptyset$. Hence, $\text{dep}(x, b) \subseteq B + b$.
  - $\text{dep}(x, b)$ is minimal tight set containing $b$. $B$ is also a tight set. Hence $\text{dep}(x, b) \cup B = B + b$ is a tight set, a contradiction.

... proof of Thm 14.5.9 continued.

- Next, assume that $x$ is found to be extreme. We need to show that we get $\text{dep}(x, e_j) = A_j$ for $e_j \in \text{sat}(x)$, with $1 \leq j \leq n = |\text{sat}(x)|$.
  - Each $A_j$ is tight since each $B$ in the algorithm is tight.
  - $e_j \in A_j$, so $\text{dep}(x, e_j) \subseteq A_j$.
  - Suppose $\text{dep}(x, e_j) \neq A_j$, and let $b$ be maximal according to $\preceq$ within $A_j \setminus \text{dep}(x, e_j)$, meaning $b \preceq c$ for any $c \in A_j \setminus \text{dep}(x, e_j)$.
  - But then $b$ is also maximal in $A_j$ since $b \notin \text{dep}(x, e_j)$ (i.e., $b \preceq c$ for any $c \in \text{dep}(x, e_j)$).
  - Hence, by Theorem 14.5.7, $A_j - b$ is also tight.
  - But if $A_j - b$ is tight, then $b$ would have been removed by line 10 of the algorithm. Hence, $b \in A_j \setminus \text{dep}(x, e_j)$ existing is a contradiction to what the algorithm does.
  - Hence, $\text{dep}(x, e_j) = A_j$.

...
Algorithm Correctness

... proof of Thm 14.5.9 continued.

- Last thing we need is to show that we get \( \text{dep}(x, e) \) for all \( e \in \text{sat}(x) \), which is equivalent to having \( B = \text{sat}(x) \) at termination.
  - \( B \) being tight means that \( B \subseteq \text{sat}(x) \), and also at termination we have \( \text{supp}(x) \subseteq B \subseteq \text{sat}(x) \).
  - \( \text{sat}(x) \) is the maximal tight set, so \( x(\text{sat}(x)) = f(\text{sat}(x)) \).
  - We saw earlier that \( f(\text{sat}(x) \setminus \text{supp}(x)|\text{supp}(x)) = 0 \) when \( x \) is extreme, which implies that \( f(A|\text{supp}(x)) = 0 \) for any \( A \subseteq \text{sat}(x) \setminus \text{supp}(x) \).
  - Of course, for any \( A \subseteq \text{sat}(x) \setminus \text{supp}(x) \), we have \( x(\text{supp}(x) + A) = x(\text{supp}(x)) \).
  - So, starting with a tight \( B \supseteq \text{supp}(x) \), we can make singleton additions to \( B \) retaining tightness, and the algorithm won’t be forced to stop doing that until it adds all of \( \text{sat}(x) \setminus \text{supp}(x) \).

---

On Greedy, and linear programming max

Theorem 14.5.10

Let \( y \in P_f \) be an extreme point, and let \( \preceq \) be the partial order of \( y \). Let \( c \in \mathbb{R}^E \). Then, \( y \) is the solution in:

\[
c^\top y = \max \{ c^\top x : x \in P_f \}
\]

iff the following three conditions hold:

1. \( c(e) \geq 0 \) for every \( e \in \text{supp}(y) \)
2. \( c(e) \leq 0 \) for every \( e \in E \setminus \text{sat}(y) \), and
3. For \( d, e \in \text{sat}(y) \) and \( d \preceq e \) imply that \( c(d) \geq c(e) \).
Separators and directed graph

- Given extreme point \( x \in P_f \), the ordering \( \preceq \) associated with \( x \) can produce a directed graph \( D = (E, F) \) where \( E \) are the vertices of the graph and \((e_1, e_2) \in F\) is a directed edge of the graph iff \( e_1 \) covers \( e_2 \) (meaning \( e_2 \preceq e_1 \)).
- A separator \( A \) of \( f \) is a set such that \( f(E) = f(A) + f(E \setminus A) \). Hence, the polytope on axes \( A, E \setminus A \) is hyperrectangular.
- The elementary separators correspond to the minimal non-empty separators, i.e., \( E_1, E_2, \ldots, E_k \) such that \( \bigcup_{i=1}^{k} E_k = E \) and \( f(E) = \sum_{i=1}^{k} f(E_k) \) and where no further refinement of this partition has this property.
- Hence, for any \( A \subseteq E \), we have
  \[
  f(A) = \sum_{i=1}^{k} f(A \cap E_k) \tag{14.29}
  \]
  So “dependence” lives only within an \( E_i \) but not between two \( E_i, E_j \) for \( i \neq j \).
- This is often important in practice (e.g., graph cut has \( |E_i| = 2 \)).

We can compute the elementary separators by constructing a directed graph for an ordering \( \preceq \). In fact:

**Theorem 14.5.11**

*Let \( x \) be an extreme point of \( P_f \) where \( x \) generated by an ordering of the entire set \( E \). Let \( D \) be the directed graph of \( \preceq \) of \( x \). Then the elementary separators of \( f \) are the vertex-sets of the connected components of \( D \).*

In order to implement an algorithm for the Theorem, we need to compute an extreme point \( x \), then the associated directed graph \( D \), and finally its connected components. \( x \) may be computed using the greedy algorithm for *any* ordering of \( E \)!
Sources for Today’s Lecture
