From Potentials to Polyhedra: Inference in Structured Models

Sebastian Nowozin

Machine Learning and Perception Group
Microsoft Research Cambridge

Colorado Springs, 20th June 2011
Approximating a Unit Disc

- Using linear inequalities, how can we approximate the unit disc?
Naive approach

- Error $\epsilon = \frac{1}{\cos \frac{\pi}{k}} - 1 \approx \frac{\pi^2}{2k^2}$
- Inefficient, $\epsilon \leq 10^{-6}$ needs $k > 2200$
- Can we do better?
Naive approach

- **Error** $\epsilon = \frac{1}{\cos \frac{\pi}{k}} - 1 \approx \frac{\pi^2}{2k^2}$
- Inefficient, $\epsilon \leq 10^{-6}$ needs $k > 2200$
- Can we do better?
Naive approach

\[
\epsilon = \frac{1}{\cos \frac{\pi}{k}} - 1 \approx \frac{\pi^2}{2k^2}
\]

- Efficient, \( \epsilon \leq 10^{-6} \) needs \( k > 2200 \)
- Can we do better?
Naive approach

- Error $\epsilon = \frac{1}{\cos \frac{\pi}{k}} - 1 \approx \frac{\pi^2}{2k^2}$
- Inefficient, $\epsilon \leq 10^{-6}$ needs $k > 2200$
- Can we do better?
Extended Formulations

- Augment variable set \((x_1, x_2)\) to \((x_1, x_2, \alpha)\)
- Define set \(S\) on enlarged space
- Project
  \[ C = \text{proj}_{x_1, x_2} S \]

- Amazing fact in high dimensions:
  Simple \(S\) (small number of inequalities) can create complicated \(C\)
  (exponential number of inequalities)
Extended Formulations

- Augment variable set \((x_1, x_2)\) to \((x_1, x_2, \alpha)\)
- Define set \(S\) on enlarged space
- Project
  \[
  C = \text{proj}_{x_1, x_2} S
  \]
- Amazing fact in high dimensions:
  Simple \(S\) (small number of inequalities) can create complicated \(C\)
  (exponential number of inequalities)
Ben-Tal/Nemirovski Polyhedron

Variables $x_1$, $x_2$, and $\alpha = (\xi^j, \eta^j)_{j=0,\ldots,k}$, parameter $k$

\begin{align*}
\xi^0 &\geq x_1, \quad \xi^0 \geq -x_1, \\
\eta^0 &\geq x_2, \quad \eta^0 \geq -x_2, \\
\xi^j &= \cos \left( \frac{\pi}{2j+1} \right) \xi^{j-1} + \sin \left( \frac{\pi}{2j+1} \right) \eta^{j-1}, \quad j = 1, \ldots, k \\
\eta^j &\geq -\sin \left( \frac{\pi}{2j+1} \right) \xi^{j-1} + \cos \left( \frac{\pi}{2j+1} \right) \eta^{j-1}, \quad j = 1, \ldots, k \\
\eta^j &\geq \sin \left( \frac{\pi}{2j+1} \right) \xi^{j-1} - \cos \left( \frac{\pi}{2j+1} \right) \eta^{j-1}, \quad j = 1, \ldots, k \\
\xi^k &\leq 1, \\
\eta^k &\leq \tan \left( \frac{\pi}{2k+1} \right) \xi^k.
\end{align*}
Ben-Tal/Nemirovski Polyhedron (cont)

Projection of Ben–Tal–Nemirovski polytope, k=2

Sebastian Nowozin
Microsoft Research Cambridge

From Potentials to Polyhedra: Inference in Structured Models
Ben-Tal/Nemirovski Polyhedron (cont)

Projection of Ben–Tal–Nemirovski polytope, \( k=3 \)
Ben-Tal/Nemirovski Polyhedron (cont)

Projection of Ben–Tal–Nemirovski polytope, $k=4$
Ben-Tal/Nemirovski Polyhedron (cont)

Projection of Ben–Tal–Nemirovski polytope, $k=7$
Ben-Tal/Nemirovski Polyhedron (cont)

- BTN-$k$, for $k = 2, 3, 4, \ldots$
- Number of non-zero coefficients in system: $9k + 11$, linear in $k$
- Number of vertices in $(x_1, x_2)$-projection: $2^{k+1}$

<table>
<thead>
<tr>
<th>$k$</th>
<th>No. vert.</th>
<th>NNZ</th>
<th>$\epsilon$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>32</td>
<td>47</td>
<td>0.0048</td>
</tr>
<tr>
<td>5</td>
<td>64</td>
<td>56</td>
<td>0.0012</td>
</tr>
<tr>
<td>6</td>
<td>128</td>
<td>65</td>
<td>$3.0 \cdot 10^{-4}$</td>
</tr>
<tr>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
</tr>
<tr>
<td>$k$</td>
<td>$2^{k+1}$</td>
<td>$9k + 11$</td>
<td>$O\left(\frac{1}{4^k}\right)$</td>
</tr>
</tbody>
</table>
Ben-Tal/Nemirovski Polyhedron (cont)

- BTN-\(k\), for \(k = 2, 3, 4, \ldots\)
- Number of non-zero coefficients in system: \(9k + 11\), linear in \(k\)
- Number of vertices in \((x_1, x_2)\)-projection: \(2^{k+1}\)
- BTN: error \(\epsilon = \frac{1}{\cos \frac{\pi}{2k+1}} - 1 = O\left(\frac{1}{4^k}\right)\) (\(\epsilon \leq 3 \cdot 10^{-7}\) for \(k = 12\))
- Naive: error \(\epsilon = \frac{1}{\cos \frac{\pi}{k}} - 1 \approx \frac{\pi^2}{2k^2}\) (\(\epsilon \leq 10^{-6}\) for \(k = 2, 200\))
- → A much better approximation
From Sets to Functions

Connections to the Literature

- Extended formulations for polyhedral sets (Balas, 1975)
- Extended formulations for convex functions in integer programs (Miller and Wolsey, 2003)

In computer vision (under various names, often combined with an inference method)

- (Rother and Kohli, 2011)
- (Ladicky et al., ECCV 2010)
- (Ishikawa, CVPR 2009)
- ...
Higher-order Interactions

Problem: graphical model formulation not expressive enough to capture structure of $E_F$,

Decomposable higher-order interactions
  - Representable by a set of $T$ new variables with state spaces $S_t$,
  - $T$, $S_t$ bounded by a polynomial in the scope size and variable state spaces
Higher-order Interactions

Problem: graphical model formulation not expressive enough to capture structure of $E_F$.

Decomposable higher-order interactions
  - Representable by a set of $T$ new variables with state spaces $S_t$,
  - $T$, $S_t$ bounded by a polynomial in the scope size and variable state spaces
Higher-order Interactions

- Problem: graphical model formulation not expressive enough to capture structure of $E_F$.
- Decomposable higher-order interactions
  - Representable by a set of $T$ new variables with state spaces $S_t$,
  - $T$, $S_t$ bounded by a polynomial in the scope size and variable state spaces
Higher-order Interactions

- Problem: graphical model formulation not expressive enough to capture structure of $E_F$,
- Decomposable higher-order interactions
  - Representable by a set of $T$ new variables with state spaces $S_t$,
  - $T$, $S_t$ bounded by a polynomial in the scope size and variable state spaces
Decomposable Higher-order Interactions

1. Partition $\mathcal{Y}_F$ into a small set $\mathcal{Z}$ of equivalence classes,
2. Introduce a new model variable $Z \in \mathcal{Z}$
3. Build simple energy model for each class (e.g. constant)
4. Integrate with original variables
Decomposable Higher-order Interactions

1. Partition $\mathcal{Y}_F$ into a small set $\mathcal{Z}$ of equivalence classes,
2. Introduce a new model variable $Z \in \mathcal{Z}$
3. Build simple energy model for each class (e.g. constant)
4. Integrate with original variables
Decomposable Higher-order Interactions

1. Partition $\mathcal{Y}_F$ into a small set $\mathcal{Z}$ of equivalence classes,
2. Introduce a new model variable $Z \in \mathcal{Z}$
3. Build simple energy model for each class (e.g. constant)
4. Integrate with original variables

Sebastian Nowozin
Microsoft Research Cambridge
From Potentials to Polyhedra: Inference in Structured Models
Example 1: Pattern-based Potential

- (Rother et al., CVPR 2009), (Komodakis and Paragios, CVPR 2009)
- Match a small set of patterns with low energy or assign a default energy
- Pattern set $\mathcal{P}$,

$$E_F(y_F) = \begin{cases} 
C_{y_F} & \text{if } y_F \in \mathcal{P} \\
C_{\max} & \text{otherwise.}
\end{cases}$$
Example 1: Pattern-based Potential (cont)

\[ E_F(y_F) = \begin{cases} 
C_{y_F} & \text{if } y_F \in \mathcal{P} \\
C_{\max} & \text{otherwise.}
\end{cases} \]

- Fix joint configuration \( y_F \)
- Pattern cost \( C_{y_F} \) or \( C_{\max} \)
Example 2: Co-occurrence Potential

- (Ladicky et al., ECCV 2010), (Delong et al., CVPR 2010)
- Have a cost function based on what sets of labels appear (independent of their counts)
Example 2: Co-occurrence Potential

\[ E_F(y) = C(L(y)) = C(\{\circ, \circ, \circ\}) \]

- (Ladicky et al., ECCV 2010), (Delong et al., CVPR 2010)
- Have a cost function based on what sets of labels appear (independent of their counts)
Example 2: Co-occurrence Potential (cont)

\[ E_F(y) = C(L(y)) = C(\{\bullet, \bullet, \bullet\}) \]

- Extended formulation with “has-color”-variable
- This extended formulation: further conditions required for \( E_F \)
- Extension possible for arbitrary \( E_F \)
- Size polynomial in the number of subsets
Example 2: Co-occurrence Potential (cont)

- Extended formulation with “has-color”-variable
- This extended formulation: further conditions required for $E_F$
- Extension possible for arbitrary $E_F$
- Size polynomial in the number of subsets
Example 2: Co-occurrence Potential (cont)

- Extended formulation with “has-color”-variable
- This extended formulation: further conditions required for $E_F$
- Extension possible for arbitrary $E_F$
- Size polynomial in the number of subsets
Non-Decomposable Interactions

Non-decomposable,
- Not representable by a small set of new variables with small state spaces
- Requires analysis outside the graphical model framework

Examples of non-decomposable interactions
- Cooperative cuts (Jegelka and Bilmes, CVPR 2011)
- Topological constraints (Vicente et al., CVPR 2008), (Nowozin and Lampert, CVPR 2009), (Chen et al., CVPR 2011)
Connectivity: Connected Subgraph Polytope

Object segmentation

- “Connectedness”: the resulting object segmentations should be connected
- (Nowozin and Lampert, CVPR 2009), (Nowozin and Lampert, SIAM IMS 2010)

Steps

- Global potential $\psi_V$: connectivity
  - Derive a polyhedral set which captures connected subgraphs
  - This set is the connected subgraph polytope
  - Use MAP-MRF linear programming relaxation, but intersect with this set
Connectivity: Connected Subgraph Polytope

Object segmentation

- “Connectedness”: the resulting object segmentations should be connected
- (Nowozin and Lampert, CVPR 2009), (Nowozin and Lampert, SIAM IMS 2010)

Steps

- Global potential $\psi_V$: connectivity
- Derive a polyhedral set which captures connected subgraphs
- This set is the *connected subgraph polytope*
- Use MAP-MRF linear programming relaxation, but *intersect* with this set
Connectivity: Connected Subgraph Polytope

Object segmentation

▶ “Connectedness”: the resulting object segmentations should be connected
▶ (Nowozin and Lampert, CVPR 2009), (Nowozin and Lampert, SIAM IMS 2010)

Steps

▶ Global potential $\psi_V$: connectivity
▶ Derive a polyhedral set which captures connected subgraphs
▶ This set is the connected subgraph polytope
▶ Use MAP-MRF linear programming relaxation, but intersect with this set
Connected Subgraph Polytope (cont)

Definition (Connected Subgraph Polytope)

Given a simple, connected, undirected graph $G = (V, E)$, consider indicator variables $y_i \in \{0, 1\}$, $i \in V$. Let $C = \{ y : G' = (V', E') \text{ connected}, \text{ with } V' = \{ i : y_i = 1 \}, E' = (V' \times V') \cap E \}$ denote the finite set of connected subgraphs of $G$. Then we call the convex hull $Z = \text{conv}(C)$ the connected subgraph polytope.
Connected Subgraph Polytope (cont)

Definition (Connected Subgraph Polytope)

Given a simple, connected, undirected graph $G = (V, E)$, consider indicator variables $y_i \in \{0, 1\}, \ i \in V$. Let $C = \{y : G' = (V', E') \text{ connected}, \ V' = \{i : y_i = 1\}, \ E' = (V' \times V') \cap E\}$ denote the finite set of connected subgraphs of $G$. Then we call the convex hull $Z = \text{conv}(C)$ the connected subgraph polytope.
Connected Subgraph Polytope (cont)

Definition (Connected Subgraph Polytope)

Given a simple, connected, undirected graph $G = (V, E)$, consider indicator variables $y_i \in \{0, 1\}$, $i \in V$. Let $C = \{y : G' = (V', E') \text{ connected, with } V' = \{i : y_i = 1\}, E' = (V' \times V') \cap E\}$ denote the finite set of connected subgraphs of $G$. Then we call the convex hull $Z = \text{conv}(C)$ the connected subgraph polytope.
Connected Subgraph Polytope (cont)

Definition (Connected Subgraph Polytope)

Given a simple, connected, undirected graph \( G = (V, E) \), consider indicator variables \( y_i \in \{0, 1\}, \ i \in V \). Let \( C = \{y : G' = (V', E') \text{ connected}, \ V' = \{i : y_i = 1\}, \ E' = (V' \times V') \cap E\} \) denote the finite set of connected subgraphs of \( G \). Then we call the convex hull \( Z = \text{conv}(C) \) the connected subgraph polytope.
Definition (Connected Subgraph Polytope)

Given a simple, connected, undirected graph $G = (V, E)$, consider indicator variables $y_i \in \{0, 1\}$, $i \in V$. Let $C = \{ y : G' = (V', E') \text{ connected}, \text{ with } V' = \{ i : y_i = 1 \}, E' = (V' \times V') \cap E \}$ denote the finite set of connected subgraphs of $G$. Then we call the convex hull $Z = \text{conv}(C)$ the connected subgraph polytope.
Connected Subgraph Polytope (cont)

Definition (Connected Subgraph Polytope)

Given a simple, connected, undirected graph $G = (V, E)$, consider indicator variables $y_i \in \{0, 1\}$, $i \in V$. Let $C = \{y : G' = (V', E') \text{ connected}, \text{ with } V' = \{i : y_i = 1\}, E' = (V' \times V') \cap E\}$ denote the finite set of connected subgraphs of $G$. Then we call the convex hull $Z = \text{conv}(C)$ the connected subgraph polytope.
Facets and Valid Inequalities

Convex polytopes have two equivalent representations

- As a convex combination of extreme points
- As a set of facet-defining linear inequalities

A linear inequality with respect to a polytope can be

- valid, does not cut off the polytope,
- representing a face, valid and touching,
- facet-defining, representing a face of dimension one less than the polytope.
Warmup

Some basic properties about the connected subgraph polytope $Z$. Note that $Z$ depends on the graph structure.

**Lemma**

*If $G$ is connected, $\dim(Z) = |V|$, that is, $Z$ has full dimension.*

**Lemma**

*For all $i \in V$, the inequalities $y_i \geq 0$ and $y_i \leq 1$ are facet-defining for $Z$.***
An Exponential-sized Class of Facet-defining Inequalities

**Theorem**

The following linear inequalities are facet-defining for $Z = \text{conv}(C)$.

$$y_i + y_j - \sum_{k \in S} y_k \leq 1, \quad \forall (i,j) \notin E : \forall S \in \bar{S}(i,j).$$

(1)

$$y_0 + y_2 - y_1 \leq 1.$$
Intuition

\[ y_i + y_j - \sum_{k \in S} y_k \leq 1, \quad \forall (i, j) \notin E : \forall S \in \tilde{S}(i, j) \]

If two vertices \( i \) and \( j \) are selected \((y_i = y_j = 1, \text{shown in black})\), then any set of vertices which separate them (set \( S \)) must contain at least one selected vertex.

**Figure:** Vertex \( i \) and \( j \) and one vertex separator set \( S \in \tilde{S}(i, j) \).
Formulation

Theorem

\( C, the set of all connected subgraphs, can be described exactly by the following constraint set. \)

\[
y_i + y_j - \sum_{k \in S} y_k \leq 1, \forall (i, j) \notin E : \forall S \in S(i, j),
\]

\[
y_i \in \{0, 1\}, \quad i \in V.
\]

\( \) (Problem): number of inequalities (2) is exponential in \(|V|\).

This means

- inequalities together with integrality are a formulation of the set of connected subgraphs,
- we can attempt to relax (3) to

\[
y_i \in [0; 1], \quad i \in V.
\]
Conclusions

- Discrete graphical models are just one way to capture structure
- There are other tractable/approximable structures
  - Extended formulations (latent variables with specific tying)
  - Polyhedral combinatorics

Open questions

- How to perform probabilistic inference in higher-order models?
- How to parametrize and learn higher-order models?
- (Is there a more suitable formalism than either graphical models or polytopes?)
Conclusions

- Discrete graphical models are just one way to capture structure
- There are other tractable/approximable structures
  - Extended formulations (latent variables with specific tying)
  - Polyhedral combinatorics

Open questions

- How to perform probabilistic inference in higher-order models?
- How to parametrize and learn higher-order models?
- (Is there a more suitable formalism than either graphical models or polytopes?)
Thank you!

feedback most welcome

nowozin@gmail.com