Higher Order Models in Computer Vision: Inference

Carsten Rother, Sebastian Nowozin

Machine Learning and Perception Group
Microsoft Research Cambridge

San Francisco, 18th June 2010
Belief Propagation Reminder

- Sum-product belief propagation
- Max-product/Max-sum/Min-sum belief propagation

Uses

- **Messages**: vectors at the factor graph edges
  - \( q_{Y_i \rightarrow F} \in \mathbb{R}^{Y_i} \), the *variable-to-factor message*, and
  - \( r_{F \rightarrow Y_i} \in \mathbb{R}^{Y_i} \), the *factor-to-variable message*.

- **Updates**: iteratively recomputing these vectors
Belief Propagation Reminder: Min-Sum Updates (cont)

**Variable-to-factor message**

\[ q_{Y_i \rightarrow F}(y_i) = \sum_{F' \in M(i) \setminus \{F\}} r_{F' \rightarrow Y_i}(y_i) \]

**Factor-to-variable message**

\[ r_{F \rightarrow Y_i}(y_i) = \min_{y'_F \in Y_F} \left( E_F(y'_F) + \sum_{j \in N(F) \setminus \{i\}} q_{Y_j \rightarrow F}(y'_j) \right) \]
Belief Propagation Reminder: Min-Sum Updates (cont)

Variable-to-factor message

\[ q_{Y_i \rightarrow F}(y_i) = \sum_{F' \in M(i) \setminus \{F\}} r_{F' \rightarrow Y_i}(y_i) \]

Factor-to-variable message

\[ r_{F \rightarrow Y_i}(y_i) = \min_{y'_F \in Y_F \atop y'_i = y_i} \left( E_F(y'_F) + \sum_{j \in N(F) \setminus \{i\}} q_{Y_j \rightarrow F}(y'_j) \right) \]
Higher-order Factors

Factor-to-variable message

\[ r_{F \to Y_i}(y_i) = \min_{y'_F \in \mathcal{Y}_F, y'_i = y_i} \left( E_F(y'_F) + \sum_{j \in N(F) \setminus \{i\}} q_{Y_j \to F}(y'_j) \right) \]

- Minimization becomes intractable for general \( E_F \)
- *Tractable* for special structure of \( E_F \)
- See (Felzenszwalb and Huttenlocher, CVPR 2004), (Potetz, CVPR 2007), and (Tarlow et al., AISTATS 2010)
Higher-order Factors

Factor-to-variable message

\[ r_{F \rightarrow Y_i}(y_i) = \min_{y'_F \in \mathcal{Y}_F} \left( E_F(y'_F) + \sum_{j \in N(F) \setminus \{i\}} q_{Y_j \rightarrow F}(y'_j) \right) \]

- Minimization becomes intractable for general \( E_F \)
- *Tractable* for special structure of \( E_F \)
- See (Felzenszwalb and Huttenlocher, CVPR 2004), (Potetz, CVPR 2007), and (Tarlow et al., AISTATS 2010)
Problem Relaxations

- Optimization problems (minimizing $g : \mathcal{G} \to \mathbb{R}$ over $\mathcal{Y} \subseteq \mathcal{G}$) can become easier if
  - feasible set is enlarged, and/or
  - objective function is replaced with a bound.
Problem Relaxations

- Optimization problems (minimizing \( g : \mathcal{G} \rightarrow \mathbb{R} \) over \( \mathcal{Y} \subseteq \mathcal{G} \)) can become easier if
  - feasible set is enlarged, and/or
  - objective function is replaced with a bound.

Definition (Relaxation (Geoffrion, 1974))

Given two optimization problems \((g, \mathcal{Y}, \mathcal{G})\) and \((h, \mathcal{Z}, \mathcal{G})\), the problem \((h, \mathcal{Z}, \mathcal{G})\) is said to be a relaxation of \((g, \mathcal{Y}, \mathcal{G})\) if,

1. \( \mathcal{Z} \supseteq \mathcal{Y} \), i.e. the feasible set of the relaxation contains the feasible set of the original problem, and
2. \( \forall y \in \mathcal{Y} : h(y) \leq g(y) \), i.e. over the original feasible set the objective function \( h \) achieves no larger values than the objective function \( g \).
   (Minimization version)
Relaxations

- Relaxed solution $z^*$ provides a bound:
  
  $h(z^*) \geq g(y^*)$, (maximization: upper bound)
  
  $h(z^*) \leq g(y^*)$, (minimization: lower bound)

- Relaxation is typically tractable

- Evidence that relaxations are “better” for learning (Kulesza and Pereira, 2007), (Finley and Joachims, 2008), (Martins et al., 2009)

- Drawback: $z^* \in Z \setminus Y$ possible
  
  - solution is not integral (but fractional), or
  
  - solution violates constraints (primal infeasible).

- To obtain $z' \in Y$ from $y^*$: rounding, heuristics
Relaxations

- Relaxed solution $z^*$ provides a bound:
  
  $$ h(z^*) \geq g(y^*), \text{ (maximization: upper bound)} $$
  $$ h(z^*) \leq g(y^*), \text{ (minimization: lower bound)} $$

- Relaxation is typically tractable

- Evidence that relaxations are “better” for learning (Kulesza and Pereira, 2007), (Finley and Joachims, 2008), (Martins et al., 2009)

- Drawback: $z^* \in \mathcal{Z} \setminus \mathcal{Y}$ possible
  - solution is not integral (but fractional), or
  - solution violates constraints (primal infeasible).

- To obtain $z' \in \mathcal{Y}$ from $y^*$: rounding, heuristics
Constructing Relaxations

Principled methods to *construct* relaxations

- Linear Programming Relaxations
- Lagrangian relaxation
- Lagrangian/Dual decomposition
Constructing Relaxations

Principled methods to *construct* relaxations

- **Linear Programming Relaxations**
- Lagrangian relaxation
- Lagrangian/Dual decomposition
MAP-MRF Linear Programming Relaxation

Simple two variable pairwise MRF

\[ Y_1 \times Y_2 \]

\[ \theta_{1,2} \]
MAP-MRF Linear Programming Relaxation

- Energy $E(y; x) = \theta_1(y_1; x) + \theta_{1,2}(y_1, y_2; x) + \theta_2(y_2; x)$
- Probability $p(y|x) = \frac{1}{Z(x)} \exp\{-E(y; x)\}$
- MAP prediction: $\arg\max_{y \in \mathcal{Y}} p(y|x) = \arg\min_{y \in \mathcal{Y}} E(y; x)$

$y_1 = 2$ $y_2 = 3$ $(y_1, y_2) = (2, 3)$
MAP-MRF Linear Programming Relaxation

\[ \mu_1 \in \{0, 1\}^{\mathcal{Y}_1} \]
\[ \mu_{1,2} \in \{0, 1\}^{\mathcal{Y}_1 \times \mathcal{Y}_2} \]
\[ \mu_2 \in \{0, 1\}^{\mathcal{Y}_2} \]

\[ \mu_1(y_1) = \sum_{y_2 \in \mathcal{Y}_2} \mu_{1,2}(y_1, y_2) \]
\[ \sum_{y_1 \in \mathcal{Y}_1} \mu_1(y_1) = 1 \]
\[ \sum_{(y_1, y_2) \in \mathcal{Y}_1 \times \mathcal{Y}_2} \mu_{1,2}(y_1, y_2) = 1 \]
\[ \sum_{y_2 \in \mathcal{Y}_2} \mu_2(y_2) = 1 \]

Energy is now linear: \( E(y; x) = \langle \theta, \mu \rangle \)
MAP-MRF Linear Programming Relaxation (cont)

\[
\begin{align*}
\min_{\mu} \quad & \sum_{i \in V} \sum_{y_i \in \mathcal{Y}_i} \theta_i(y_i) \mu_i(y_i) + \sum_{\{i,j\} \in E} \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \theta_{i,j}(y_i, y_j) \mu_{i,j}(y_i, y_j) \\
\text{s.t.} \quad & \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V, \\
& \sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i \\
& \mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i, \\
& \mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j.
\end{align*}
\]
MAP-MRF Linear Programming Relaxation (cont)

\[
\begin{align*}
\min_{\mu} & \quad \sum_{i \in V} \sum_{y_i \in \mathcal{Y}_i} \theta_i(y_i) \mu_i(y_i) + \sum_{\{i,j\} \in E} \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \theta_{i,j}(y_i, y_j) \mu_{i,j}(y_i, y_j) \\
\text{s.t.} & \quad \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V, \\
& \quad \sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i,j\} \in E, \forall y_i \in \mathcal{Y}_i \\
& \quad \mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i, \\
& \quad \mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i,j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j.
\end{align*}
\]

\[\Rightarrow\] NP-hard integer linear program
MAP-MRF Linear Programming Relaxation (cont)

\[
\begin{align*}
\min_{\mu} \quad & \sum_{i \in V} \sum_{y_i \in \mathcal{Y}_i} \theta_i(y_i) \mu_i(y_i) + \sum_{\{i,j\} \in E} \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \theta_{i,j}(y_i, y_j) \mu_{i,j}(y_i, y_j) \\
\text{s.t.} \quad & \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V, \\
& \sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i \\
& \mu_i(y_i) \in [0, 1], \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i, \\
& \mu_{i,j}(y_i, y_j) \in [0, 1], \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j.
\end{align*}
\]

▶ → linear program
▶ → LOCAL polytope
MAP-MRF LP Relaxation, Works

MAP-MRF LP Analysis

- (Wainwright et al., Trans. Inf. Tech., 2005), (Weiss et al., UAI 2007), (Werner, PAMI 2005)
- TRW-S (Kolmogorov, PAMI 2006)

Improving tightness

- (Komodakis and Paragios, ECCV 2008), (Kumar and Torr, ICML 2008), (Sontag and Jaakkola, NIPS 2007), (Sontag et al., UAI 2008), (Werner, CVPR 2008), (Kumar et al., NIPS 2007)

Algorithms based on the LP

- (Globerson and Jaakkola, NIPS 2007), (Kumar and Torr, ICML 2008), (Sontag et al., UAI 2008)
MAP-MRF LP Relaxation, General case

\[
\sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1,
\sum_{y_A \in \mathcal{Y}_A} [y_A]_B = y_B \mu_A(y_A) = \mu_B(y_B), \quad \forall y_B \in \mathcal{Y}_B.
\]

Generalization to higher-order factors

- For $B \subseteq A$ factors: marginalization constraints
- Defines LOCAL polytope (Wainwright and Jordan, 2008)
MAP-MRF LP Relaxation, Known results

1. LOCAL is tight iff the factor graph is a forest (acyclic)
2. All labelings are vertices of LOCAL (Wainwright and Jordan, 2008)
3. For cyclic graphs there are additional fractional vertices.
4. If all factors have regular energies, the fractional solutions are never optimal (Wainwright and Jordan, 2008)
5. For models with only binary states: half-integrality, integral variables are certain

See some examples later...
Constructing Relaxations

Principled methods to *construct* relaxations

- Linear Programming Relaxations
- **Lagrangian relaxation**
- Lagrangian/Dual decomposition
Lagrangian Relaxation, Motivation

\[
\sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = \mu_B(y_B), \quad \forall y_B \in \mathcal{Y}_B.
\]

MAP-MRF LP with higher-order factors problematic:

- Number of LP variables grows with \(|\mathcal{Y}_A| \rightarrow \text{exponential in MRF variables}
- But: number of constraints stays small (\(|\mathcal{Y}_B|\))

*Lagrangian Relaxation* allows one to solve the LP by treating the constraints implicitly.
Lagrangian Relaxation

$$\min_y g(y)$$
$$\text{sb.t.} \quad y \in \mathcal{D},$$
$$\quad y \in \mathcal{C}.$$ 

Assumption

- Optimizing $g(y)$ over $y \in \mathcal{D}$ is “easy”.
- Optimizing over $y \in \mathcal{D} \cap \mathcal{C}$ is hard.

High-level idea

- Incorporate $y \in \mathcal{C}$ constraint into objective function

Lagrangian Relaxation (cont)

Recipe

1. Express $C$ in terms of equalities and inequalities

$$C = \{ y \in \mathcal{G} : u_i(y) = 0, \forall i = 1, \ldots, I, \ v_j(y) \leq 0, \forall j = 1, \ldots, J \},$$

- $u_i : \mathcal{G} \to \mathbb{R}$ differentiable, typically affine,
- $v_j : \mathcal{G} \to \mathbb{R}$ differentiable, typically convex.

2. Introduce Lagrange multipliers, yielding

$$\min_y g(y) \\text{s.t.} \ y \in \mathcal{D},$$
$$u_i(y) = 0 : \lambda, \ i = 1, \ldots, I,$$
$$v_j(y) \leq 0 : \mu, \ j = 1, \ldots, J.$$
Lagrangian Relaxation (cont)

2. Introduce Lagrange multipliers, yielding

$$\min_y g(y) \quad \text{s.t.} \quad y \in D, \quad u_i(y) = 0 : \lambda, \quad i = 1, \ldots, l, \quad v_j(y) \leq 0 : \mu, \quad j = 1, \ldots, J.$$ 

3. Build partial Lagrangian

$$\min_y g(y) + \lambda^T u(y) + \mu^T v(y) \quad \text{(1)} \quad \text{s.t.} \quad y \in D.$$
Lagrangian Relaxation (cont)

3. Build partial Lagrangian

\[
\min_y g(y) + \lambda^T u(y) + \mu^T v(y)
\]

\[
\text{s.t.} \quad y \in \mathcal{D}.
\]

Theorem (Weak Duality of Lagrangian Relaxation)

For differentiable functions \( u_i : \mathcal{G} \to \mathbb{R} \) and \( v_j : \mathcal{G} \to \mathbb{R} \), and for any \( \lambda \in \mathbb{R}^I \) and any non-negative \( \mu \in \mathbb{R}^J, \mu \geq 0 \), problem (1) is a relaxation of the original problem: its value is lower than or equal to the optimal value of the original problem.
3. Build partial Lagrangian

\[ q(\lambda, \mu) := \min_y g(y) + \lambda^T u(y) + \mu^T v(y) \quad (1) \]
\[ \text{sb.t.} \quad y \in D. \]

4. Maximize lower bound wrt \( \lambda, \mu \)

\[ \max_{\lambda,\mu} q(\lambda, \mu) \]
\[ \text{sb.t.} \quad \mu \geq 0 \]

\[ \rightarrow \text{efficiently solvable if } q(\lambda, \mu) \text{ can be evaluated} \]
Optimizing Lagrangian Dual Functions

4. Maximize lower bound wrt $\lambda, \mu$

$$\max_{\lambda, \mu} q(\lambda, \mu)$$ (2)

$$\text{s.t.} \quad \mu \geq 0$$

**Theorem (Lagrangian Dual Function)**

1. $q$ is concave in $\lambda$ and $\mu$, Problem (2) is a concave maximization problem.

2. If $q$ is unbounded above, then the original problem is infeasible.

3. For any $\lambda, \mu \geq 0$, let

$$y(\lambda, \mu) = \arg\min_{y \in \mathcal{D}} g(y) + \lambda^T u(y) + \mu^T v(u)$$

Then, a supergradient of $q$ can be constructed by evaluating the constraint functions at $y(\lambda, \mu)$ as

$$u(y(\lambda, \mu)) \in \frac{\partial}{\partial \lambda} q(\lambda, \mu), \quad \text{and} \quad v(y(\lambda, \mu)) \in \frac{\partial}{\partial \mu} q(\lambda, \mu).$$
Subgradients, Supergradients

- Subgradients for convex functions: global affine underestimators
- Supergradients for concave functions: global affine overestimators
Subgradients, Supergradients

- Subgradients for convex functions: global affine underestimators
- Supergradients for concave functions: global affine overestimators
Subgradients, Supergradients

Subgradients for convex functions: global affine underestimators

Supergradients for concave functions: global affine overestimators

\[ f(y) \geq f(x) + g^\top (y - x), \forall y \]
Subgradients, Supergradients

- Subgradients for convex functions: global affine underestimators
- Supergradients for concave functions: global affine overestimators
Example behaviour of objectives

(from a 10-by-10 pairwise binary MRF relaxation, submodular)
Primal Solution Recovery

Assume we solved the dual problem for \((\lambda^*, \mu^*)\)

1. Can we obtain a primal solution \(y^*\)?
2. Can we say something about the relaxation quality?

**Theorem (Sufficient Optimality Conditions)**

If for a given \(\lambda, \mu \geq 0\), we have \(u(y(\lambda, \mu)) = 0\) and \(v(y(\lambda, \mu)) \leq 0\) (primal feasibility) and further we have

\[
\lambda^\top u(y(\lambda, \mu)) = 0, \quad \text{and} \quad \mu^\top v(y(\lambda, \mu)) = 0,
\]

(complementary slackness), then

- \(y(\lambda, \mu)\) is an optimal primal solution to the original problem, and
- \((\lambda, \mu)\) is an optimal solution to the dual problem.
Assume we solved the dual problem for \((\lambda^*, \mu^*)\)

1. Can we obtain a primal solution \(y^*\)?
2. Can we say something about the relaxation quality?

**Theorem (Sufficient Optimality Conditions)**

*If for a given \(\lambda, \mu \geq 0\), we have \(u(y(\lambda, \mu)) = 0\) and \(v(y(\lambda, \mu)) \leq 0\) (primal feasibility) and further we have

\[
\lambda^\top u(y(\lambda, \mu)) = 0, \quad \text{and} \quad \mu^\top v(y(\lambda, \mu)) = 0,
\]

(complementary slackness), then

- \(y(\lambda, \mu)\) is an optimal primal solution to the original problem, and
- \((\lambda, \mu)\) is an optimal solution to the dual problem.*
Primal Solution Recovery (cont)

But,

- we might never see a solution satisfying the optimality condition
- is the case only if there is no duality gap, i.e. \( q(\lambda^*, \mu^*) = g(x, y^*) \)

Special case: integer linear programs

- we can always reconstruct a primal solution to

\[
\min_y \quad g(y) \\
\text{s.t.} \quad y \in \text{conv}(D), \quad y \in \mathcal{C}.
\]  

- For example for subgradient method updates it is known that (Anstreicher and Wolsey, MathProg 2009)

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} y(\lambda^t, \mu^t) \to y^* \text{ of (3)}. 
\]
Application to higher-order factors

\[ \sum_{y_A \in Y_A} \mu_A(y_A) = \mu_B(y_B), \quad \forall y_B \in \mathcal{Y}_B. \]

- (Komodakis and Paragios, CVPR 2009) propose a MAP inference algorithm suitable for higher-order factors.
- Can be seen as Lagrangian relaxation of the marginalization constraint of the MAP-MRF LP.
- See also (Johnson et al., ACCC 2007), (Schlesinger and Giginyak, CSC 2007), (Wainwright and Jordan, 2008, Sect. 4.1.3), (Werner, PAMI 2005), (Werner, TechReport 2009), (Werner, CVPR 2008).
Komodakis MAP-MRF Decomposition

\[
\min_{\mu} \sum_{i \in V} \sum_{y_i \in \mathcal{Y}_i} \theta_i(y_i) \mu_i(y_i) + \sum_{\{i,j\} \in E} \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \theta_{i,j}(y_i, y_j) \mu_{i,j}(y_i, y_j)
\]

s.t. \[
\sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V,
\]
\[
\sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E,
\]
\[
\sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i
\]
\[
\mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i,
\]
\[
\mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j.
\]
Komodakis MAP-MRF Decomposition

\[
\min_{\mu} \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle \\
\text{s.b.t.} \quad \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V, \\
\sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E, \\
\sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i \\
\mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i, \\
\mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j.
\]
**Komodakis MAP-MRF Decomposition**

\[
\begin{align*}
\min_{\mu} \quad & \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle + \sum_{y_A \in \mathcal{Y}_A} \theta_A(y_A) \mu_A(y_A) \\
\text{s.t.} \quad & \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V, \\
& \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E, \\
& \sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i \\
& \sum_{y_A \in \mathcal{Y}_A, [y_A]_B = y_B} \mu_A(y_A) = \mu_B(y_B), \quad \forall B \subset A, y_B \in \mathcal{Y}_B, \\
& \sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1, \\
& \mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i, \\
& \mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j, \\
& \mu_A(y_A) \in \{0, 1\}, \quad \forall y_A \in \mathcal{Y}_A.
\end{align*}
\]
Komodakis MAP-MRF Decomposition

\[
\begin{align*}
\min_{\mu} & \quad \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle + \sum_{y_A \in Y_A} \theta_A(y_A) \mu_A(y_A) \\
\text{s.t.} & \quad \sum_{y_i \in Y_i} \mu_i(y_i) = 1, \quad \forall i \in V, \\
& \quad \sum_{(y_i, y_j) \in Y_i \times Y_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E, \\
& \quad \sum_{y_j \in Y_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in Y_i \\
& \quad \sum_{y_A \in Y_A, [y_A]_B = y_B} \mu_A(y_A) = \mu_B(y_B) : \phi_A \rightarrow_B (y_B), \quad \forall B \subset A, y_B \in Y_B, \\
& \quad \sum_{y_A \in Y_A} \mu_A(y_A) = 1, \\
& \quad \mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in Y_i, \\
& \quad \mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in Y_i \times Y_j, \\
& \quad \mu_A(y_A) \in \{0, 1\}, \quad \forall y_A \in Y_A.
\end{align*}
\]
Komodakis MAP-MRF Decomposition

\[
q(\phi) := \min_{\mu} \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle + \sum_{y_A \in \mathcal{Y}_A} \theta_A(y_A) \mu_A(y_A)
\]

\[
+ \sum_{B \subset A, \ y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \left( \sum_{y_A \in \mathcal{Y}_A, [y_A]_B = y_B} \mu_A(y_A) - \mu_B(y_B) \right)
\]

s.t.
\[
\sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V,
\]
\[
\sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E,
\]
\[
\sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1,
\]
\[
\sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i
\]
\[
\mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i,
\]
\[
\mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j
\]
\[
\mu_i(y_i) \in \{0, 1\}, \quad \forall y_i \in \mathcal{Y}_i
\]
Komodakis MAP-MRF Decomposition

\[ q(\phi) := \min_{\mu} \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle + \sum_{y_A \in \mathcal{Y}_A} \theta_A(y_A) \mu_A(y_A) \]

\[ + \sum_{B \subset A, \ y_B \in \mathcal{Y}_B} \phi_{A \to B}(y_B) \sum_{y_A \in \mathcal{Y}_A, [y_A]_B = y_B} \mu_A(y_A) - \sum_{B \subset A, \ y_B \in \mathcal{Y}_B} \phi_{A \to B}(y_B) \mu_B(y_B) \]

s.t. \[ \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \ \forall i \in V, \]
\[ \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \ \forall \{i, j\} \in E, \]
\[ \sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1, \]
\[ \sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \ \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i \]
\[ \mu_i(y_i) \in \{0, 1\}, \ \forall i \in V, \forall y_i \in \mathcal{Y}_i, \]
\[ \mu_{i,j}(y_i, y_j) \in \{0, 1\}, \ \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j, \]
\[ \mu_A(y_A) \in \{0, 1\}, \ \forall y_A \in \mathcal{Y}_A. \]
Komodakis MAP-MRF Decomposition

\[ q(\phi) := \min_{\mu} \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle - \sum_{B \subseteq A, y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \mu_B(y_B) + \sum_{B \subseteq A, y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \sum_{y_A \in \mathcal{Y}_A, [y_A]_B = y_B} \mu_A(y_A) + \sum_{y_A \in \mathcal{Y}_A} \theta_A(y_A) \mu_A(y_A) \]

s.t. \[ \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V, \]

\[ \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E, \]

\[ \sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1, \]

\[ \sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i \]

\[ \mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i, \]

\[ \mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j, \]

\[ \mu_A(y_A) \in \{0, 1\}, \quad \forall y_A \in \mathcal{Y}_A. \]
Komodakis MAP-MRF Decomposition

\[ q(\phi) := \min_\mu \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle - \sum_{B \subset A, \ y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \mu_B(y_B) \]

\[ + \sum_{B \subset A, \ y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \sum_{y_A \in \mathcal{Y}_A, [y_A]_B = y_B} \mu_A(y_A) + \sum_{y_A \in \mathcal{Y}_A} \theta_A(y_A) \mu_A(y_A) \]

s.t. \[ \sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V, \]

\[ \sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E, \]

\[ \sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1, \]

\[ \sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i \]

\[ \mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i, \]

\[ \mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j, \]

\[ \mu_A(y_A) \in \{0, 1\}, \quad \forall y_A \in \mathcal{Y}_A. \]
Komodakis MAP-MRF Decomposition

\[
q(\phi) := \min_\mu \sum_{i \in V} \langle \theta_i, \mu_i \rangle + \sum_{\{i,j\} \in E} \langle \theta_{i,j}, \mu_{i,j} \rangle - \sum_{B \subseteq A, \ y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \mu_B(y_B) + \sum_{B \subseteq A, \ y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \sum_{y_A \in \mathcal{Y}_A, [y_A]_B = y_B} \mu_A(y_A) + \sum_{y_A \in \mathcal{Y}_A} \theta_A(y_A) \mu_A(y_A)
\]

sb.t.

\[
\sum_{y_i \in \mathcal{Y}_i} \mu_i(y_i) = 1, \quad \forall i \in V,
\]

\[
\sum_{(y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = 1, \quad \forall \{i, j\} \in E,
\]

\[
\sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1,
\]

\[
\sum_{y_j \in \mathcal{Y}_j} \mu_{i,j}(y_i, y_j) = \mu_i(y_i), \quad \forall \{i, j\} \in E, \forall y_i \in \mathcal{Y}_i
\]

\[
\mu_i(y_i) \in \{0, 1\}, \quad \forall i \in V, \forall y_i \in \mathcal{Y}_i,
\]

\[
\mu_{i,j}(y_i, y_j) \in \{0, 1\}, \quad \forall \{i, j\} \in E, \forall (y_i, y_j) \in \mathcal{Y}_i \times \mathcal{Y}_j,
\]

\[
\mu_A(y_A) \in \{0, 1\}, \quad \forall y_A \in \mathcal{Y}_A.
\]
Maximizing $q(\phi)$ corresponds to message passing from the higher-order factor to the unary factors.
The higher-order subproblem

$$\min_{\mu_A} \sum_{B \subseteq A, \ y_B \in \mathcal{Y}_B} \phi_{A \rightarrow B}(y_B) \sum_{y_A \in \mathcal{Y}_A, [y_A]_B = y_B} \mu_A(y_A) + \sum_{y_A \in \mathcal{Y}_A} \theta_A(y_A) \mu_A(y_A)$$

s.t. \( \sum_{y_A \in \mathcal{Y}_A} \mu_A(y_A) = 1, \)

\( \mu_A(y_A) \in \{0, 1\}, \ \forall y_A \in \mathcal{Y}_A. \)

- Simple structure: select one element
- Simplify to an optimization problem over \( y_A \in \mathcal{Y}_A \)
The higher-order subproblem

\[
\min_{y_A \in \mathcal{Y}_A} \theta_A(y_A) + \sum_{B \subset A, y_B \in \mathcal{Y}_B} I([y_A]_B = y_B) \phi_{A \rightarrow B}(y_B)
\]

- Solvability depends on structure of $\theta_A$ as function of $y_A$
- (For now) assume we can solve for $y_A^* \in \mathcal{Y}_A$
- How can we use this to solve the original large LP?
Maximizing the dual function: Subgradient Method

Maximizing $q(\phi)$:

1. Initialize $\phi$
2. Iterate
   2.1 Given $\phi$ solve pairwise MAP subproblem, obtain $\mu_1^*$, $\mu_2^*$
   2.2 Given $\phi$ solve higher-order MAP subproblem, obtain $y_A^*$
   2.3 Update $\phi$
      ▶ Supergradient

\[
\frac{\partial}{\partial \phi_B(y_B)} \equiv \sum_{B \subset A, \ y_B \in Y_B} (I([y_A^*]_B = y_B) - \mu_B(y_B))
\]

▶ Supergradient ascent, step size $\alpha_t > 0$

\[
\phi_B(y_B) \leftarrow \phi_B(y_B) + \alpha_t \sum_{B \subset A, \ y_B \in Y_B} (I([y_A^*]_B = y_B) - \mu_B(y_B))
\]

The step size $\alpha_t$ must satisfy the “diminishing step size conditions”,
i) $\lim_{t \to \infty} \alpha_t = 0$, and ii) $\sum_{t=0}^{\infty} \alpha_t = \infty$. 
Subgradient Method (cont)

Typical step size choices

1. Simple diminishing step size,

\[ \alpha^t = \frac{1 + m}{t + m}, \]

where \( m > 0 \) is an arbitrary constant.

2. Polyak’s step size,

\[ \alpha^t = \beta^t \frac{\bar{q} - q(\lambda^t, \mu^t)}{\|u(y(\lambda^t, \mu^t))\|^2 + \|v(y(\lambda^t, \mu^t))\|^2}, \]

where

- \( 0 < \beta^t < 2 \) is a diminishing step size, and
- \( \bar{q} \geq q(\lambda^*, \mu^*) \) is an upper bound on the optimal dual objective.
Subgradient Method (cont)

Primal solution recovery

- Uniform average

\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} y(\lambda^t, \mu^t) \to y^*.
\]

- Geometric series (volume algorithm)

\[
\bar{y}^t = \gamma y(\lambda^t, \mu^t) + (1 - \gamma)\bar{y}^{t-1},
\]

where \(0 < \gamma < 1\) is a small constant such as \(\gamma = 0.1\), and \(\bar{y}^t\) might be slightly primal infeasible.
Pattern-based higher order potentials

Example (Pattern-based potentials)

Given a small set of patterns $\mathcal{P} = \{y_1, y_2, \ldots, y_K\}$, define

$$\theta_A(y_A) = \begin{cases} 
C_{y_A} & \text{if } y_A \in \mathcal{P}, \\
C_{\text{max}} & \text{otherwise}.
\end{cases}$$

- (Rother et al, CVPR 2009), (Komodakis and Paragios, CVPR 2009)
- Generalizes $\mathcal{P}^n$-Potts model (Kohli et al., CVPR 2007)
- “Patterns have low energies”: $C_y \leq C_{\text{max}}$ for all $y \in \mathcal{P}$
Pattern-based higher order order potentials

Example (Pattern-based potentials)

Given a small set of patterns $\mathcal{P} = \{y_1, y_2, \ldots, y_K\}$, define

$$\theta_A(y_A) = \begin{cases} C_y & \text{if } y_A \in \mathcal{P}, \\ C_{\text{max}} & \text{otherwise}. \end{cases}$$

Can solve

$$\min_{y_A \in \mathcal{Y}_A} \theta_A(y_A) + \sum_{B \subset A, \ y_B \in \mathcal{Y}_B} I([y_A]_B = y_B) \phi_{A \rightarrow B}(y_B)$$

by

1. Testing all $y_A \in \mathcal{P}$, and
2. Fixing $\theta_A(y_A) = C_{\text{max}}$ and solving for the minimum of the second term.
Constructing Relaxations

Principled methods to *construct* relaxations

- Linear Programming Relaxations
- Lagrangian relaxation
- Lagrangian/Dual decomposition
Dual/Lagrangian Decomposition

Additive structure

\[
\min_y \sum_{k=1}^{K} g_k(y) \\
\text{sb.t.} \quad y \in \mathcal{Y}_k, \quad \forall k = 1, \ldots, K,
\]

such that \( \sum_{k=1}^{K} g_k(y) \) is hard, but for any \( k \),

\[
\min_y g_k(y) \\
\text{sb.t.} \quad y \in \mathcal{Y}_k
\]

is tractable.

- (Guignard, “Lagrangean Decomposition”, MathProg 1987)
- (Conejo et al., “Decomposition Techniques in Mathematical Programming”, 2006)
Dual/Lagrangian Decomposition

Idea

1. Selectively duplicate variables ("variable splitting")

\[
\min_{y_1, \ldots, y_K, y} \sum_{k=1}^{K} g_k(y_k)
\]

s.t. \( y \in \mathcal{Y}', \)
\( y_k \in \mathcal{Y}_k, \quad k = 1, \ldots, K, \)
\( y = y_k : \lambda_k, \quad k = 1, \ldots, K, \quad (4) \)

where \( \mathcal{Y}' \supseteq \bigcap_{k=1,\ldots,K} \mathcal{Y}_k. \)

2. Add coupling equality constraints between duplicated variables

3. Apply Lagrangian relaxation to the coupling constraint

Also known as dual decomposition and variable splitting.
Dual/Lagrangian Decomposition (cont)

\[
\begin{align*}
\min_{y_1, \ldots, y_K, y} & \quad \sum_{k=1}^{K} g_k(y_k) \\
\text{s.t.} & \quad y \in \mathcal{Y}', \\
& \quad y_k \in \mathcal{Y}_k, \quad k = 1, \ldots, K, \\
& \quad y = y_k : \lambda_k, \quad k = 1, \ldots, K.
\end{align*}
\] (5)
Dual/Lagrangian Decomposition (cont)

\[
\begin{align*}
\min_{y_1, \ldots, y_K, y} & \quad \sum_{k=1}^{K} g_k(y_k) + \sum_{k=1}^{K} \lambda_k^\top (y - y_k) \\
\text{s.t.} & \quad y \in \mathcal{Y}', \\
& \quad y_k \in \mathcal{Y}_k, \quad k = 1, \ldots, K,
\end{align*}
\]
Dual/Lagrangian Decomposition (cont)

\[
\begin{align*}
\min_{y_1, \ldots, y_K, y} & \quad \sum_{k=1}^{K} \left( g_k(y_k) - \lambda_k^\top y_k \right) + \left( \sum_{k=1}^{K} \lambda_k \right)^\top y \\
\text{s.t.} & \quad y \in \mathcal{Y}', \\
& \quad y_k \in \mathcal{Y}_k, \quad k = 1, \ldots, K,
\end{align*}
\]

- Problem is *decomposed*
- There can be multiple possible decompositions of different quality
Dual/Lagrangian Decomposition (cont)

Dual function

\[ q(\lambda) := \min_{y_1, \ldots, y_K, y} \sum_{k=1}^{K} (g_k(y_k) - \lambda_k^\top y_k) + \left(\sum_{k=1}^{K} \lambda_k\right)^\top y \]

s.t. \[ y \in \mathcal{Y}', \]
\[ y_k \in \mathcal{Y}_k, \quad k = 1, \ldots, K, \]

Dual maximization problem

\[ \max_{\lambda} \quad q(\lambda) \]

s.t. \[ \sum_{k=1}^{K} \lambda_k = 0, \]

where \( \{\lambda | \sum_{k=1}^{K} \lambda_k \neq 0\} \) is the domain where \( q(\lambda) > -\infty \).
Dual/Lagrangian Decomposition (cont)

Dual maximization problem

\[
\max_\lambda q(\lambda)
\]

s.b.t.

\[
\sum_{k=1}^{K} \lambda_k = 0,
\]

Projected subgradient method, using

\[
\frac{\partial}{\partial \lambda_k} \ni (y - y_k) - \frac{1}{K} \sum_{\ell=1}^{K} (y - y_\ell) = \frac{1}{K} \sum_{\ell=1}^{K} y_\ell - y_k.
\]
Example (Connectivity, Vicente et al., CVPR 2008)
Example (Connectivity, Vicente et al., CVPR 2008)
Example (Connectivity, Vicente et al., CVPR 2008)

\[
\min_y E(y) + C(y),
\]

where \( E(y) \) is a normal binary pairwise segmentation energy, and

\[
C(y) = \begin{cases} 
0 & \text{if marked points are connected,} \\
\infty & \text{otherwise.}
\end{cases}
\]
Example (Connectivity, Vicente et al., CVPR 2008)

\[
\begin{align*}
\min_{y,z} & \quad E(y) + C(z) \\
\text{s.t.} & \quad y = z.
\end{align*}
\]
Example (Connectivity, Vicente et al., CVPR 2008)

\[
\min_{y,z} \quad E(y) + C(z)
\]
\[
\text{sa.t.} \quad y = z : \lambda.
\]
Example (Connectivity, Vicente et al., CVPR 2008)

\[
\min_{y,z} \quad E(y) + C(z) + \lambda^T (y - z)
\]
Example (Connectivity, Vicente et al., CVPR 2008)

\[ q(\lambda) := \min_{y,z} E(y) + \lambda^\top y + C(z) - \lambda^\top z \]

- **Subproblem 1**: normal binary image segmentation energy
- **Subproblem 2**: shortest path problem (no pairwise terms)
- Maximize \( q(\lambda) \) to find best lower bound
- Above derivation simplified, see (Vicente et al., CVPR 2008) for a more sophisticated decomposition using three subproblems
Example (Connectivity, Vicente et al., CVPR 2008)

\[ q(\lambda) \,:= \, \min_{y,z} \left( E(y) + \lambda^\top y + C(z) - \lambda^\top z \right) \]

- **Subproblem 1**: normal binary image segmentation energy
- **Subproblem 2**: shortest path problem (no pairwise terms)
- \( \rightarrow \) Maximize \( q(\lambda) \) to find best lower bound
- Above derivation simplified, see (Vicente et al., CVPR 2008) for a more sophisticated decomposition using three subproblems
Other examples

Plenty of applications of decomposition:

- (Komodakis et al., ICCV 2007)
- (Strandmark and Kahl, CVPR 2010)
- (Torresani et al., ECCV 2008)
- (Werner, TechReport 2009)
- (Strandmark and Kahl, CVPR 2010)
- (Kumar and Koller, CVPR 2010)
- (Woodford et al., ICCV 2009)
- (Vicente et al., ICCV 2009)
Geometric Interpretation

**Theorem (Lagrangian Decomposition Primal)**

Let $\mathcal{Y}_k$ be finite sets and $g(y) = c^\top y$ a linear objective function. Then the optimal solution of the Lagrangian decomposition obtains the value of the following relaxed primal optimization problem.

$$
\min_y \sum_{k=1}^{K} g_k(y) \\
\text{s.t. } y \in \text{conv}(\mathcal{Y}_k), \quad \forall k = 1, \ldots, K.
$$

Therefore, optimizing the Lagrangian decomposition dual is equivalent to

- optimizing the primal objective,
- on the intersection of the convex hulls of the individual feasible sets.
Geometric Interpretation

**Theorem (Lagrangian Decomposition Primal)**

Let $\mathcal{Y}_k$ be finite sets and $g(y) = c^\top y$ a linear objective function. Then the optimal solution of the Lagrangian decomposition obtains the value of the following relaxed primal optimization problem.

\[
\min_y \sum_{k=1}^{K} g_k(y) \\
\text{s.t. } y \in \text{conv}(\mathcal{Y}_k), \quad \forall k = 1, \ldots, K.
\]

Therefore, optimizing the Lagrangian decomposition dual is equivalent to

- optimizing the primal objective,
- on the intersection of the convex hulls of the individual feasible sets.
Geometric Interpretation (cont)
Geometric Interpretation (cont)
Geometric Interpretation (cont)
## Geometric Interpretation (cont)

Consider the geometric interpretation of the optimization problem:

- **Convex Sets:**
  - \( \text{conv}(\mathcal{V}_1) \)
  - \( \text{conv}(\mathcal{V}_2) \)
  - \( \text{conv}(\mathcal{V}_1) \cap \text{conv}(\mathcal{V}_2) \)

- **Points:**
  - \( y_D \)
  - \( y^* \)

- **Relations:**
  - \( c^T y \)
  - \( \text{conv}(\mathcal{V}_2) \)

The problem is to find the point \( y^* \) that lies in the intersection of \( \text{conv}(\mathcal{V}_1) \cap \text{conv}(\mathcal{V}_2) \) that maximizes \( c^T y \).
Lagrangian Relaxation vs. Decomposition

Lagrangian Relaxation
- Needs *explicit* complicating constraints

Lagrangian/Dual Decomposition
- Can work with black-box solver for subproblem (*implicit*)
- Can yield stronger bounds
Lagrangian Relaxation vs. Decomposition

\[ \mathcal{V}_1' \supseteq \text{conv} (\mathcal{V}_1) \]

\[ \text{conv} (\mathcal{V}_1) \cap \text{conv} (\mathcal{V}_2) \]

\[ \text{conv} (\mathcal{V}_2) \]

\[ y^* \]

\[ y_R \]

\[ y_D \]

\[ c^\top y \]
Polyhedral View on Higher-order Interactions

Idea

- MAP-MRF LP relaxation: LOCAL polytope
- Incorporate higher-order interactions directly by constraining LOCAL
- Techniques from *polyhedral combinatorics*
- See (Werner, CVPR 2008), (Nowozin and Lampert, CVPR 2009), (Lempitsky et al., ICCV 2009)
Representation of Polytopes

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.
Intersecting Polytopes

Construction

1. Consider *marginal polytope* $\mathcal{M}(G)$ of a discrete graphical model
2. Vertices of $\mathcal{M}(G)$ are feasible labelings
3. Add linear inequalities $d^T \mu \leq b$ that cut off vertices
4. Optimize over resulting intersection $\mathcal{M}'(G)$,

$$E_A(\mu_A) = \begin{cases} 0 & \text{if } \mu \text{ is valid,} \\ \infty & \text{otherwise.} \end{cases}$$
Intersecting Polytopes

Construction

1. Consider *marginal polytope* $\mathcal{M}(G)$ of a discrete graphical model
2. Vertices of $\mathcal{M}(G)$ are feasible labelings
3. Add linear inequalities $d^T \mu \leq b$ that cut off vertices
4. Optimize over resulting intersection $\mathcal{M}'(G)$,

$$E_A(\mu_A) = \begin{cases} 0 & \text{if } \mu \text{ is valid}, \\ \infty & \text{otherwise.} \end{cases}$$
Intersecting Polytopes

Construction

1. Consider *marginal polytope* \( \mathcal{M}(G) \) of a discrete graphical model
2. Vertices of \( \mathcal{M}(G) \) are feasible labelings
3. Add linear inequalities \( d^T \mu \leq b \) that cut off vertices
4. Optimize over resulting intersection \( \mathcal{M}'(G) \),

\[
E_A(\mu_A) = \begin{cases} 
0 & \text{if } \mu \text{ is valid,} \\
\infty & \text{otherwise.}
\end{cases}
\]
Intersecting Polytopes

Construction

1. Consider *marginal polytope* $\mathcal{M}(G)$ of a discrete graphical model
2. Vertices of $\mathcal{M}(G)$ are feasible labelings
3. Add linear inequalities $d^T \mu \leq b$ that cut off vertices
4. Optimize over resulting intersection $\mathcal{M}'(G)$,

$$E_A(\mu_A) = \begin{cases} 0 & \text{if } \mu \text{ is valid,} \\ \infty & \text{otherwise.} \end{cases}$$
Defining the Inequalities

Questions

1. Where do the inequalities come from?
2. Is this construction always possible?
3. How to optimize over $\mathcal{M}'(G)$?
1. Deriving Combinatorial Polytopes

- “Foreground mask must be connected”
- Global constraint on labelings
1. Deriving Combinatorial Polytopes

- “Foreground mask must be connected”
- Global constraint on labelings
Constructing the Inequalities
Constructing the Inequalities
Constructing the Inequalities
Constructing the Inequalities
Constructing the Inequalities
Constructing the Inequalities

\[ y_0 + y_2 - y_1 \leq 1 \]
Generalization and Intuition

The following inequalities are violated by unconnected subgraphs:

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

*Intuition:* “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 \(S \in \bar{S}(i, j)\).
Derivation, Formal Steps
Steps from intuition to inequalities:

1. Identify invariants that hold for any “good” solution
2. Formulate invariants as linear inequalities
3. Prove that inequalities are facet-defining
4. Prove that together with integrality they are a formulation

Final result is a statement like:

Theorem
For a given graph $G = (V, E)$, the set of all induced subgraphs that are connected, can be described exactly by the following constraint set.

\[
y_i + y_j - \sum_{k \in S} y_k \leq 1, \forall (i, j) \notin E : \forall S \in S(i, j),
\]
\[
y_i \in \{0, 1\}, \quad i \in V.
\]
Derivation, Formal Steps

Steps from intuition to inequalities:

1. Identify invariants that hold for any “good” solution
2. Formulate invariants as linear inequalities
3. Prove that inequalities are facet-defining
4. Prove that together with integrality they are a formulation

Final result is a statement like:

Theorem

For a given graph $G = (V, E)$, the set of all induced subgraphs that are connected, can be described exactly by the following constraint set.

$$y_i + y_j - \sum_{k \in S} y_k \leq 1, \forall (i, j) \notin E : \forall S \in S(i, j),$$ $$y_i \in \{0, 1\}, \quad i \in V.$$
2. Is this always possible?

Yes, but...

- identifying and proving the inequalities is not straight-forward,
- even if all facet-defining inequalities are known, problem can remain intractable.

Further reading

- Polyhedral combinatorics literature (Schrijver, 2003), (Ziegler, 1994)
- There exist software to help identifying facet-defining inequalities, see e.g. polymake
3. How to optimize over $\mathcal{M}'(G)$?

- LOCAL has a small number of inequalities,
- $\mathcal{M}'(G)$ typically has an exponential number of inequalities,
- Optimization over $\mathcal{M}'(G)$ is still tractable if separation problem can be solved.

**Definition (Separation problem)**

Given a point $y$, prove that $y \in \mathcal{M}'(G)$ or produce a valid inequality that is strictly violated.
Example Separation for Connectivity

Need to show that for all \((i, j) \notin E\) with \(y_i > 0, y_j > 0\):

\[
y_i + y_j - \sum_{k \in S} y_k - 1 \leq 0, \quad \forall S \in \bar{S}(i, j).
\]

Equivalently, in variational form:

\[
\max_{S \in \bar{S}(i, j)} \left( y_i + y_j - \sum_{k \in S} y_k - 1 \right) \leq 0. \tag{5}
\]

If (5) is satisfied, no violated inequality for \((i, j)\) exist. Otherwise, the set corresponding to the violated inequality is

\[
S^*(i, j) = \arg\min_{S \in \bar{S}(i, j)} \sum_{k \in S} y_k. \tag{6}
\]
Example Separation for Connectivity (cont)

Need to solve

\[ S^*(i, j) = \arg\min_{S \subseteq \bar{S}(i, j)} \sum_{k \in S} y_k \]

- Transform \( G \) into a directed edge-weighted graph
- Solve minimum \((i, j)\)-cut problem
- Equivalently: solve a linear max-flow problem (efficient)

It follows that,

- by construction, a finite-capacity \((i, j)\)-cut always exist,
- the cut found corresponds to a violated inequality or proves absence.
Example

X pattern

Noisy X pattern
Example

Figure: MRF/MRFcomp/CMRF: $E = -985.61, E = -974.16, E = -984.21$

Figure: MRF/MRFcomp/CMRF: $E = -980.13, E = -974.03, E = -976.83$
Bounding-box Prior

Example (Bounding Box Prior (Lempitsky et al., ICCV 2009))

- Prior: segmentation should be tight wrt bounding box
- Exponential number of linear inequalities
Relaxation and Decomposition: Summary

Lagrangian relaxation

- requires: complicating constraints
- widely applicable

Problem decomposition

- requires: multiple tractable subproblems
- can make use of efficient algorithms for the subproblems
- widely applicable

Polyhedral higher-order interactions

- requires: combinatorial understanding of constraint set
- universal: can represent any constraint
- tractability depends on tractable separation oracle