Maximum Persistency via Iterative Relaxed Inference with Graphical Models

Alexander Shekhovtsov, TU Graz
Paul Swoboda, Heidelberg University
Bogdan Savchynskyy, TU Dresden

May 27, 2015

Outline

Introduction

Persistency

ILP	LP
$\min c^\mathsf{T} x$	$\min c^\mathsf{T} x$
$Ax \leq b$	$Ax \leq b$
$x \in \{0, 1\}^n$	$x \in [0, 1]^n$

1. When the solution to LP is integer?

Persistency

ILP	LP
$\min c^{\mathsf{T}}x$	$\min c^{\mathsf{T}}x$
$Ax \leq b$	$Ax \leq b$
$x \in \{0,1\}^n$	$x \in [0,1]^n$

1. When the solution to LP is integer?

$$x = (0, 1, 1, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{3}, 1, 0, \dots)$$

2. Is the integer part of an optimal solution to LP optimal for ILP?

Persistency

ILP	LP
$\min c^\mathsf{T} x$	$\min c^\mathsf{T} x$
$Ax \leq b$	$Ax \leq b$
$x \in \{0,1\}^n$	$x \in [0,1]^n$

1. When the solution to LP is integer?

$$x = (0, 1, 1, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{3}, 1, 0, \ldots)$$

2. Is the integer part of an optimal solution to LP optimal for ILP?
3. Is any part of an optimal solution to LP optimal for ILP?

Persistency

ILP	LP
$\min c^\mathsf{T} x$	$\min c^\mathsf{T} x$
$Ax \leq b$	$Ax \leq b$
$x \in \{0,1\}^n$	$x \in [0,1]^n$

1. When the solution to LP is integer?

$$x = (0, 1, 1, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{3}, 1, 0, \dots)$$

2. Is the integer part of an optimal solution to LP optimal for ILP?
3. Is any part of an optimal solution to LP optimal for ILP?
4. Sufficient conditions for a part of an optimal solution to LP to be optimal for ILP?

Persistency

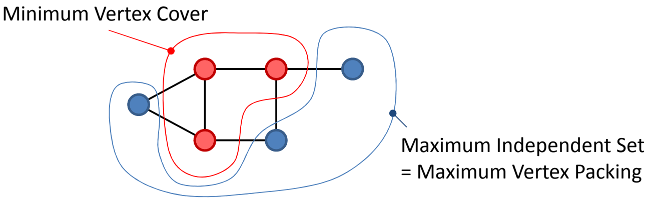
ILP	LP
$\min c^{\mathsf{T}} x$	$\min c^{\mathsf{T}} x$
$Ax \leq b$	$Ax \leq b$
$x \in \{0,1\}^n$	$x \in [0,1]^n$

1. When the solution to LP is integer?

$$x = (0, 1, 1, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{3}, 1, 0, \dots)$$

2. Is the integer part of an optimal solution to LP optimal for ILP?
3. Is any part of an optimal solution to LP optimal for ILP?
4. Sufficient conditions for a part of an optimal solution to LP to be optimal for ILP?
5. Find the largest part of LP solution satisfying 4.

Vertex Packing / Maximum Independent Set



Maximum Weighted Vertex Packing

- $(\mathcal{V}, \mathcal{E})$ – an undirected graph;
- *Vertex Packing* is a subset $P \subset \mathcal{V}$ for which $u, v \in P \Rightarrow (u, v) \notin \mathcal{E}$;
- Weights $c \colon \mathcal{V} \to \mathbb{R}$;
- Problem:

$$\max_x \sum_{v \in \mathcal{V}} c_v x_v \tag{VP}$$

$$(\forall uv \in \mathcal{E})\ x_u + x_v \leq 1,$$

$$(\forall v \in \mathcal{V})\ x_v \in \{0, 1\}.$$

Vertex Packing / Maximum Independent Set

Relaxing the integrality constraints:

$$\max_{\mu} \sum_{v \in \mathcal{V}} c_v \mu_v \qquad \text{(VPL)}$$
$$(\forall uv \in \mathcal{E})\ \mu_u + \mu_v \leq 1,$$
$$(\forall v \in \mathcal{V})\ \mu_v \geq 0.$$

Theorems

- (Balinski, 1965; Lorentzen, 1966): Any basic feasible solution to (VLP) is $\{0, \frac{1}{2}, 1\}$-valued.
- (Edmonds and Pulleyblank) (VLP) reduces to a maxflow problem on a related symmetric bipartite graph;
- (Nemhauser and Trotter, 1975): Variables which assume binary values in an optimum (VLP) solution retain the same values in an optimum (VP) solution.
- (Picard and Queyranne, 1977): There exists a unique maximum set of variables that are integer valued in an optimal solution to (VLP).

QPBO

Quadratic pseudo-Boolean Optimization (QPBO)

- $(\mathcal{V}, \mathcal{E})$ – an undirected graph;
- Weights $a\colon \mathcal{V} \cup \mathcal{E} \to \mathbb{R}$;
- Problem:

$$\min_x \sum_{v \in \mathcal{V}} a_v x_v + \sum_{uv \in \mathcal{E}} a_{uv} x_u x_v$$

$$(\forall v \in \mathcal{V})\ x_v \in \{0, 1\}.$$

- Generalizes Vertex Packing (let $a_{uv} = B$, a big number; $a = -c_v$).

QPBO

Natural linear relaxation: $x_s \to \mu_s \in [0, 1]$, $x_s x_t \to \mu_{st} \in [0, 1]$;

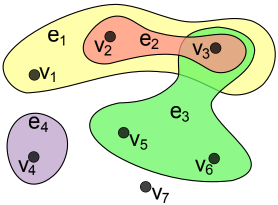
$$\min_{\mu\colon \mathcal{V}\cup\mathcal{E}\to[0,1]} \sum_{v\in\mathcal{V}} a_v \mu_v + \sum_{uv\in\mathcal{E}} a_{uv}\mu_{uv} \quad \text{s.t.} \tag{LP}$$

$$(\forall uv \in \mathcal{E})\ \mu_u + \mu_v - 1 \le \mu_{uv} \le \min(\mu_u, \mu_v).$$

Theorems

- (?): Each extreme point of the feasible set is $\{0, \frac{1}{2}, 1\}$-valued.
- (Hammer et al., 1984; Boros et al., 1991): LP reduces to a maxflow problem;
- **Weak Persistency** (Hammer et al., 1984): Variables μ_v which assume binary values in an optimum (LP) solution retain the same values in an ILP solution.
- **Strong Persistency** (Hammer et al., 1984): Variables μ_v which assume binary values in **all** optimal (LP) solutions retain the same values in **all** optimal ILP solutions.

0-1 Polynomial Programming



A hypergraph (courtesy of wikipedia).

0-1 Polynomial Programming / pseudo-Boolean Optimization

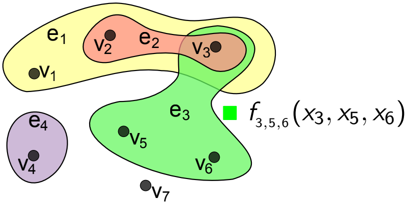
- $(\mathcal{V}, \mathcal{E})$ – a hypergraph, $\mathcal{E} \subset 2^{\mathcal{V}}$;
- Weights $f\colon \mathcal{E} \to \mathbb{R}$;
- Problem: $$\min_{x \in \{0,1\}^{\mathcal{V}}} \sum_{C \in \mathcal{E}} f_C \prod_{v \in C} x_v. \qquad \text{(PP)}$$

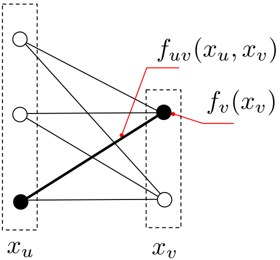
- Any pseudo-Boolean function can be represented as a multilinear polynomial.

0-1 Polynomial Programming

- Relaxation of Sherali and Adams (1990)
  - – optimal solutions are not half-integral in general;
  - – no combinatorial method to solve;
  - – does not possess persistency in general;
  - + Tightest "local" relaxation
- (Bi)submodular relaxations (Kolmogorov, 2012)
  - + extreme feasible solutions are half-integral;
  - + reduces to sum of (bi)submodular functions minimization;
  - + possess persistency;
  - – Weaker than relaxation of Sherali and Adams (1990);

Energy Minimization





Energy Minimization / Weighted Constraint Satisfaction

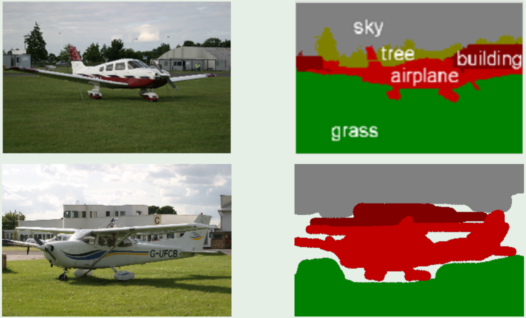
- $(\mathcal{V}, \mathcal{E})$ - a hypergraph;
- $\mathcal{X}_v$ - a finite set of *labels*, $v \in \mathcal{V}$;
- Costs $f_{\mathrm{C}}\colon \prod_{v \in \mathrm{C}} \mathcal{X}_v \to \mathbb{R}$; Cost vector $f \in \mathbb{R}^{\mathcal{I}}$.
- Energy: $E_f(x) = \sum_{\mathrm{C} \in \mathcal{E}} f_{\mathrm{C}}(x_{\mathrm{C}})$;
- Probloem: $\min_{x \in \mathcal{X}} E_f(x)$;

Energy Minimization

Example: Potts Model for Object Class Segmentation

- $\mathcal{V}$ - set of pixels; $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ neighboring pixels;
- $\mathcal{X}_s = \{1, \dots K\}$ – class label;
- $E_f(x) = \sum_{s\in\mathcal{V}} f_s(x_s) + \sum_{st\in\mathcal{E}} \lambda_{st}[\![x_s \neq x_t]\!]$.

Image Ground Truth



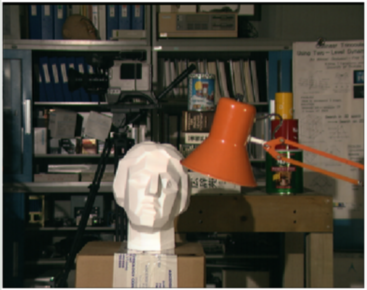
(MSRC object class segmentation)

Persistncy / Partial Optimality

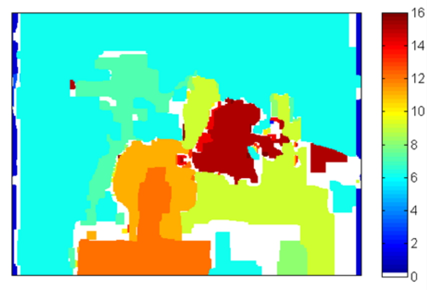
Example: Potts Model for Stereo

- $\mathcal{V}$ - set of pixels; $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ neighboring pixels;
- $\mathcal{X}_s = \{1, \dots K\}$ – disparity value;
- $E_f(x) = \sum_{s\in\mathcal{V}} f_s(x_s) + \sum_{st\in\mathcal{E}} \lambda_{st}[\![x_s \neq x_t]\!]$.

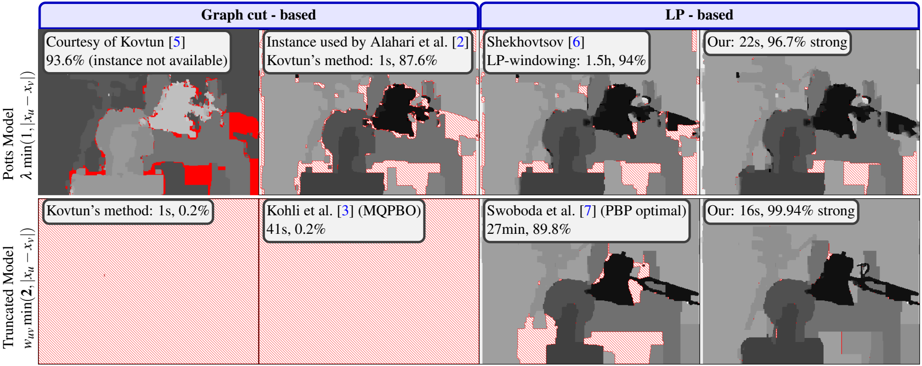
Stereo Reconstruction



Partial Optimality (Method of Kovtun (2003))



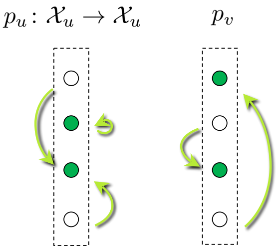
Development of Partial Optimality Methods



- Model 1 (Kovtun'03, Alahari et al.'10): Potts, strong unaries with window aggregation
- Model 2 (Szeliski et al., 2008): Nearly Potts, per-pixel unaries

Maximum Persistency Problem

Improving Mapping (substitution of labels)

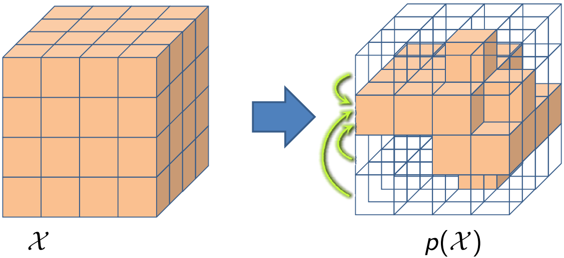


$p\colon \mathcal{X} \to \mathcal{X}$ node-wise

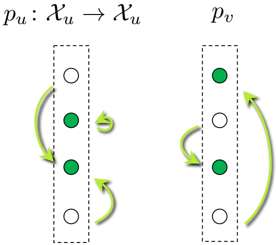
Definition

Mapping p is improving if $\forall x \; E_f(p(x)) \leq E_f(x)$.

- If x is optimal then $p(x)$ is optimal. Search space can be reduced.



Improving Mapping (substitution of labels)



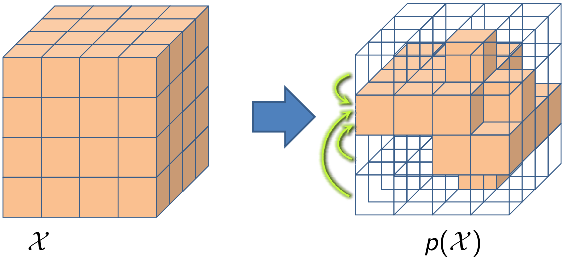
$p \colon \mathcal{X} \to \mathcal{X}$ node-wise

Definition

Mapping p is improving if $\forall x \ E_f(p(x)) \leq E_f(x)$.

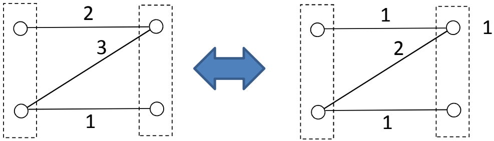
Remark: no distinction between strict/non-strict in this talk!

- If x is optimal then $p(x)$ is optimal. Search space can be reduced.



Equivalent Transformations

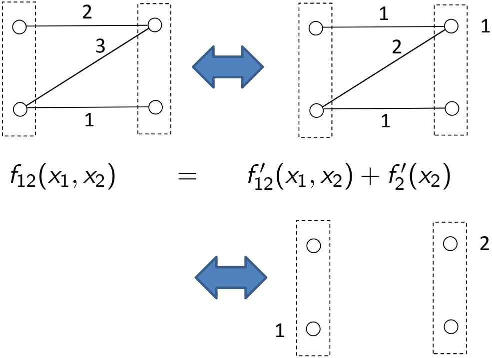
$E_f(x) = \sum_{\mathrm{C}\in\mathcal{E}} f_{\mathrm{C}}(x_{\mathrm{C}})$. Cost vectors $f, g \in \mathbb{R}^{\mathcal{I}}$ equivalent when $E_f = E_g$.



$$f_{12}(x_1, x_2) \quad = \quad f'_{12}(x_1, x_2) + f'_2(x_2)$$

Equivalent Transformations

$E_f(x) = \sum_{\mathrm{C}\in\mathcal{E}} f_{\mathrm{C}}(x_{\mathrm{C}})$. Cost vectors $f, g \in \mathbb{R}^{\mathcal{I}}$ equivalent when $E_f = E_g$.

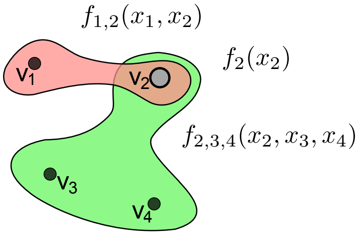


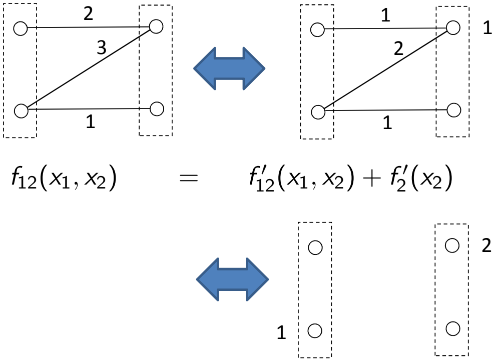
$$f_{12}(x_1, x_2) \quad = \quad f'_{12}(x_1, x_2) + f'_2(x_2)$$

$$= \quad f''_1(x_1) + 0 + f''_2(x_2)$$

Equivalent Transformations

$E_f(x) = \sum_{c \in \mathcal{E}} f_c(x_c)$. Cost vectors $f, g \in \mathbb{R}^{\mathcal{I}}$ equivalent when $E_f = E_g$.





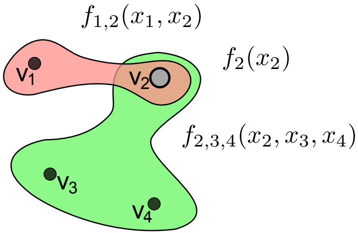
$$f_{12}(x_1, x_2) = f'_{12}(x_1, x_2) + f'_2(x_2)$$

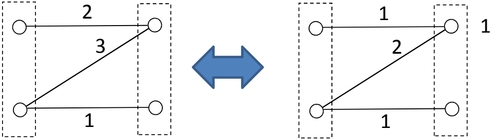
$$= f''_1(x_1) + 0 + f''_2(x_2)$$

- Affine space: $g \sim h \sim f$
  $\Rightarrow \quad E_{\alpha g + (1-\alpha)h} = E_f$.
- $f \sim g \quad \Leftrightarrow \quad (\exists \varphi)\ g = f - A^{\mathsf{T}} \varphi$.

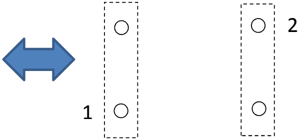
Equivalent Transformations

$E_f(x) = \sum_{\mathrm{C}\in\mathcal{E}} f_{\mathrm{C}}(x_{\mathrm{C}})$. Cost vectors $f, g \in \mathbb{R}^{\mathcal{I}}$ equivalent when $E_f = E_g$.





$f_{12}(x_1, x_2) \quad = \quad f'_{12}(x_1, x_2) + f'_2(x_2)$



$= \quad f''_1(x_1) + 0 + f''_2(x_2)$

- Affine space: $g \sim h \sim f$
  $\Rightarrow \quad E_{\alpha g+(1-\alpha)h} = E_f$.
- $f \sim g \quad \Leftrightarrow \quad (\exists\varphi)\ g = f - A^{\mathsf{T}}\varphi$.

- Equivalent Transformations, Shlezinger (1976)
- Reparametrization, Wainwright et al. (2003)
- Equivalence Preserving Transformations in WCSP

Sufficient Conditions for Peristency (Dual)

- Improving mapping: $(\forall x \in \mathcal{X})\ E_f(p(x)) \leq E_f(x)$ – NP hard to verify

Sufficient Conditions for Peristency (Dual)

- Improving mapping: $(\forall x \in \mathcal{X})\ E_f(p(x)) \leq E_f(x)$ – NP hard to verify
- Locally improving: $(\forall \mathrm{C} \in \mathcal{E}, \forall x_\mathrm{C} \in \mathcal{X}_\mathrm{C})\ f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq f_\mathrm{C}(x_\mathrm{C})$ – easy to verify.

Sufficient Conditions for Peristency (Dual)

- Improving mapping: $(\forall x \in \mathcal{X})\ E_f(p(x)) \leq E_f(x)$ – NP hard to verify
- Locally improving: $(\forall \mathrm{C} \in \mathcal{E}, \forall x_\mathrm{C} \in \mathcal{X}_\mathrm{C})\ f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq f_\mathrm{C}(x_\mathrm{C})$

  sufficient for $\sum_{\mathrm{C} \in \mathcal{E}} f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq \sum_{\mathrm{C} \in \mathcal{E}} f_\mathrm{C}(x_\mathrm{C})$

Sufficient Conditions for Peristency (Dual)

- Improving mapping: $(\forall x \in \mathcal{X})\ E_f(p(x)) \leq E_f(x)$ – NP hard to verify
- Locally improving: $(\forall \mathrm{C} \in \mathcal{E}, \forall x_\mathrm{C} \in \mathcal{X}_\mathrm{C})\ f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq f_\mathrm{C}(x_\mathrm{C})$

  sufficient for $\sum_{\mathrm{C}\in\mathcal{E}} f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq \sum_{\mathrm{C}\in\mathcal{E}} f_\mathrm{C}(x_\mathrm{C})$
- Equivalent Transformations + Locally Improving:
  $f^\varphi := f - A^\mathsf{T}\varphi$

Sufficient Conditions for Peristency (Dual)

- Improving mapping: $(\forall x \in \mathcal{X})\ E_f(p(x)) \leq E_f(x)$ – NP hard to verify
- Locally improving: $(\forall \mathrm{C} \in \mathcal{E}, \forall x_\mathrm{C} \in \mathcal{X}_\mathrm{C})\ f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq f_\mathrm{C}(x_\mathrm{C})$

  sufficient for $\sum_{\mathrm{C}\in\mathcal{E}} f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq \sum_{\mathrm{C}\in\mathcal{E}} f_\mathrm{C}(x_\mathrm{C})$
- Equivalent Transformations + Locally Improving:
  $f^\varphi := f - A^\mathsf{T}\varphi$

Definition

Mapping $p\colon X \to \mathcal{X}$ is *relaxed-improving* if
$(\exists\varphi)\ (\forall \mathrm{C} \in \mathcal{E}, \forall x_\mathrm{C} \in \mathcal{X}_\mathrm{C})\ f^\varphi_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq f^\varphi_\mathrm{C}(x_\mathrm{C})$;

A system of linear inequalities in φ.

Sufficient Conditions for Peristency (Dual)

- Improving mapping: $(\forall x \in \mathcal{X})\ E_f(p(x)) \leq E_f(x)$ – NP hard to verify
- Locally improving: $(\forall \mathrm{C} \in \mathcal{E}, \forall x_\mathrm{C} \in \mathcal{X}_\mathrm{C})\ f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq f_\mathrm{C}(x_\mathrm{C})$

  sufficient for $\displaystyle\sum_{\mathrm{C}\in\mathcal{E}} f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq \sum_{\mathrm{C}\in\mathcal{E}} f_\mathrm{C}(x_\mathrm{C})$
- Equivalent Transformations + Locally Improving:
  $f^\varphi := f - A^\mathsf{T}\varphi$

Definition

Mapping $p\colon X \to \mathcal{X}$ is *relaxed-improving* if
$(\exists\varphi)\ (\forall \mathrm{C} \in \mathcal{E}, \forall x_\mathrm{C} \in \mathcal{X}_\mathrm{C})\ f^\varphi_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C})) \leq f^\varphi_\mathrm{C}(x_\mathrm{C})$;

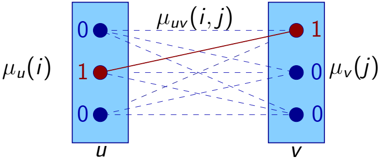
A system of linear inequalities in φ.
In matrix form: $(\exists\varphi)\ P^\mathsf{T} f^\varphi \leq f^\varphi$

$$(\exists\varphi)\ (I - P^\mathsf{T})(f - A^\mathsf{T}\varphi) \geq 0 \text{ – Linear program in } \varphi.$$

P^T: $(P^\mathsf{T} f)_\mathrm{C}(x_\mathrm{C}) = f_\mathrm{C}(p_\mathrm{C}(x_\mathrm{C}))$.

LP Relaxation

- Energy: $E_f(x) = f_\varnothing + \sum_{v\in\mathcal{V}} f_v(x_v) + \dots$
- Linearize: $E_f(x) = \langle f, \mu \rangle$
  $(f_\varnothing \mu_\varnothing + \sum_v \sum_{i\in\mathcal{X}_u} f_v(i)\mu_v(i) + \dots)$
- $\mu = \delta(x) \in \mathbb{R}^{\mathcal{I}}$ - indicator vector of costs selected by x.



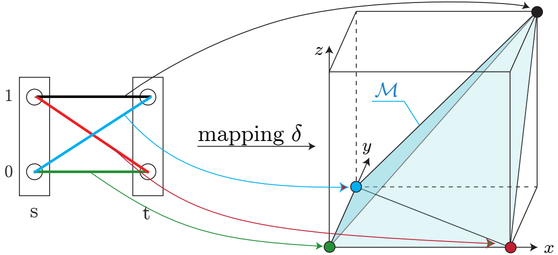
- Constraints:
  $\mu \in \{0,1\}^{\mathcal{I}}$
  $\mu_\varnothing = 1$
  $\sum_i \mu_u(i) = 1$ – normalization
  $\sum_i \mu_{uv}(i,j) = \mu_v(j)$ – marginalization
- LP relaxation: $\min \langle f, \mu \rangle$

$$A\mu = 0$$
$$\mu_\varnothing = 1$$
$$\mu \geq 0$$

LP Relaxation

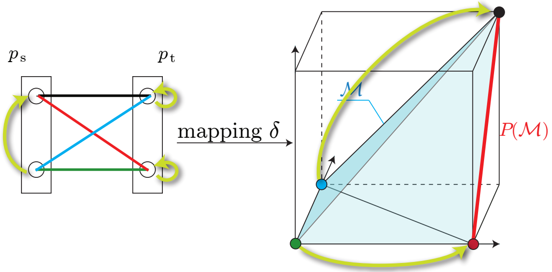
On a higher level:

- Embedding: $\delta\colon \mathcal{X} \to \mathbb{R}^{\mathcal{I}}$;
- Linearization: $E_f(x) = \langle f, \delta(x) \rangle$;
- $\min_x E_f(x) \geq \min_{\mu \in \Lambda} \langle f, \mu \rangle$;
- Polytope: $\Lambda = \{\mu \mid A\mu = 0,\ \mu_\varnothing = 1,\ \mu \geq 0\} \supset \delta(\mathcal{X})$;



Relaxed Improving Mapping

- Mapping $p\colon \mathcal{X} \to \mathcal{X}$ can be represented in the space $\mathbb{R}^{\mathcal{I}}$:



Definition

$P\colon \mathbb{R}^{\mathcal{I}} \to \mathbb{R}^{\mathcal{I}}$ is a *linear extension* of $p\colon \mathcal{X} \to \mathcal{X}$ if
$$(\forall x \in \mathcal{X})\ \delta(p(x)) = P\delta(x)$$

- An oblique projection in $\mathbb{R}^{\mathcal{I}}$.

Sufficient Conditions for Persistency (Primal)

- Recall: $p\colon \mathcal{X} \to \mathcal{X}$ is improving if $(\forall x \in \mathcal{X}) \;\; E_f(p(x)) \leq E(x)$
- In the embedding: $(\forall \mu \in \delta(\mathcal{X})) \;\; \langle f, P\mu \rangle \leq \langle f, \mu \rangle$;

Definition

Mapping $p\colon \mathcal{X} \to \mathcal{X}$ is *relaxed-improving* if

$$(\forall \mu \in \Lambda) \;\; \langle f, P\mu \rangle \leq \langle f, \mu \rangle;$$

Sufficient Conditions for Persistency (Primal)

- Recall: $p\colon \mathcal{X} \to \mathcal{X}$ is improving if $(\forall x \in \mathcal{X})\ \ E_f(p(x)) \leq E(x)$
- In the embedding: $(\forall \mu \in \delta(\mathcal{X}))\ \ \langle f, P\mu \rangle \leq \langle f, \mu \rangle$;

Definition

Mapping $p\colon \mathcal{X} \to \mathcal{X}$ is *relaxed-improving* if

$$(\forall \mu \in \Lambda)\ \ \langle f, P\mu \rangle \leq \langle f, \mu \rangle;$$

- A sufficient condition since $\Lambda \supset \delta(\mathcal{X})$

Sufficient Conditions for Persistency (Primal)

- Recall: $p \colon \mathcal{X} \to \mathcal{X}$ is improving if $(\forall x \in \mathcal{X}) \;\; E_f(p(x)) \leq E(x)$
- In the embedding: $(\forall \mu \in \delta(\mathcal{X})) \;\; \langle f, P\mu \rangle \leq \langle f, \mu \rangle$;

Definition

Mapping $p \colon \mathcal{X} \to \mathcal{X}$ is *relaxed-improving* if

$$(\forall \mu \in \Lambda) \;\; \langle f, P\mu \rangle \leq \langle f, \mu \rangle;$$

- A sufficient condition since $\Lambda \supset \delta(\mathcal{X})$
- Linear program in μ: $(\forall \mu \in \Lambda) \; \langle f, (I - P)\mu \rangle \geq 0$;

$$\min_{\mu \in \Lambda} \langle f, (I - P)\mu \rangle \geq 0.$$

Sufficient Conditions for Persistency (Primal)

- Recall: $p\colon \mathcal{X} \to \mathcal{X}$ is improving if $(\forall x \in \mathcal{X}) \ \ E_f(p(x)) \leq E(x)$
- In the embedding: $(\forall \mu \in \delta(\mathcal{X})) \ \ \langle f, P\mu\rangle \leq \langle f, \mu\rangle$;

Definition

Mapping $p\colon \mathcal{X} \to \mathcal{X}$ is *relaxed-improving* if

$$(\forall \mu \in \Lambda) \ \ \langle f, P\mu\rangle \leq \langle f, \mu\rangle;$$

- A sufficient condition since $\Lambda \supset \delta(\mathcal{X})$
- Linear program in μ: $(\forall \mu \in \Lambda)\ \langle f, (I-P)\mu\rangle \geq 0$;

$$\min_{\mu\in\Lambda}\langle f, (I-P)\mu\rangle \geq 0.$$

Theorem

Primal and Dual sufficient conditions are equivalent.

(by LP strong duality and substitution)

Generality of Sufficient Conditions

Theorems (Shekhovtsov (2014, 2015))

Relaxed-improving condition with natural (local) relaxations are satisfied for a.o.f.:

pairwise multilabel	Simple DEE (Goldstein, 1994)	✓
	MQPBO (Kohli et al., 2008)	✓
	Kovtun (2003) one-agains-all	✓
	Kovtun (2011) iterative	✓
	Swoboda et al. (2014)*	✓
	Roof dual / QPBO Hammer et al. (1984)	✓
higher order pseudo-Boolean	Reductions: HOCR (Ishikawa, 2011), (Fix et al., 2011)	FLP
	Bisubmodular relaxations (Kolmogorov, 2010)**	BLP
	Generalized Roof Dualilty (Kahl and Strandmark, 2011)	FLP
	Persistency by Adams et al. (1998)	FLP

BLP = Basic LP Relaxation Werner (2007); Thapper and Živný (2013);
FLP = Full Local LP Relaxation, equivalent to Sherali and Adams (1990);

*Swoboda et al. (2014) is higher order but the comparison proof is for pairwise case. **Result holds for sum of bisubmodular functions over the same hypergraph as the BLP relaxation.

Maximum Persistency

- Given that verification problem is polynomially solvable,
- which method is better?

Proposition

Pose "the best partial optimality" as optimization problem

Maximum Persistency

- Given that verification problem is polynomially solvable,
- which method is better?

Proposition

Pose "the best partial optimality" as optimization problem

Maximum Persistency Problem

Find the mapping $p\colon \mathcal{X} \to \mathcal{X}$ that delivers the maximum problem reduction:

$$\min_{p \in \mathcal{P}} \sum_{u \in \mathcal{V}} |p(\mathcal{X}_u)| \quad \text{s.t. } p \text{ is relaxed-improving,}$$

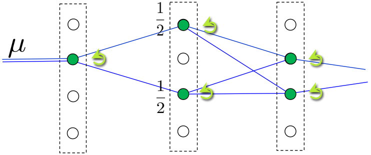
$\mathcal{P}$ - class of mappings.

The Algorithm

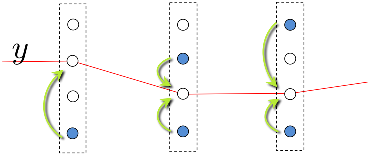
Restricted Class of Mappings

Theorem

Let μ be a solution to LP-relaxation: $\mu \in \mathrm{argmin}_{\mu\in\Lambda}\langle f, \mu\rangle$ and $p\colon \mathcal{X} \to \mathcal{X}$ be (strictly) relaxed-improving. Then $P\mu = \mu$.



- Fix a *test* labeling y from μ and try substitute other labels with it.

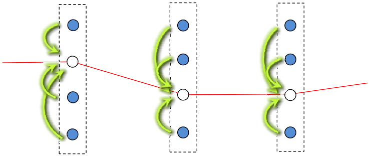


Restricted Class of Mappings

Still Covers

pairwise multilabel	Simple DEE (Goldstein, 1994)	
	MQPBO (Kohli et al., 2008)	
	Kovtun (2003) one-agains-all	
	Kovtun (2011) iterative	
	Swoboda et al. (2014)*	
higher order pseudo-Boolean	Roof dual / QPBO Hammer et al. (1984)	
	Reductions: HOCR (Ishikawa, 2011), (Fix et al., 2011)	FLP
	Bisubmodular relaxations (Kolmogorov, 2010)**	BLP
	Generalized Roof Dualilty (Kahl and Strandmark, 2011)	FLP
	Persistency by Adams et al. (1998)	FLP

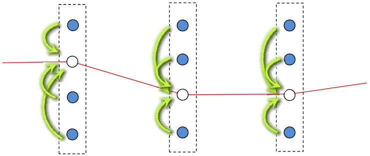
Discrete Cutting Plane



Algorithm

- Start with a mapping p that substitutes *everything* with y

Discrete Cutting Plane



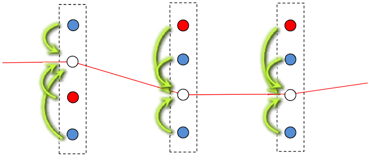
Algorithm

- Start with a mapping p that substitutes *everything* with y
- Auxiliary problem $g = (I - P^{\mathsf{T}})f$
- Check relaxed-improving conditions by solving LP dual:

$$\max_{\varphi} g^{\varphi}_{\varnothing} \quad \overset{?}{\geq} \quad 0$$

$$(\forall \mathrm{C} \in \mathcal{E},\ \forall x_{\mathrm{C}})\ g^{\varphi}_{\mathrm{C}}(x_{\mathrm{C}}) \geq 0$$

Discrete Cutting Plane



Algorithm

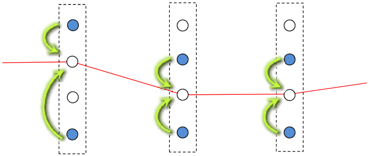
- Start with a mapping p that substitutes *everything* with y
- Auxiliary problem $g = (I - P^{\mathsf{T}})f$
- Check relaxed-improving conditions by solving LP dual:

$$\max_{\varphi} g^{\varphi}_{\varnothing} \overset{?}{\geq} 0$$

$$(\forall \mathrm{C} \in \mathcal{E},\ \forall x_{\mathrm{C}})\ g^{\varphi}_{\mathrm{C}}(x_{\mathrm{C}}) \geq 0$$

- If not satisfied, determine active (blocking) constraints on nodes

Discrete Cutting Plane



Algorithm

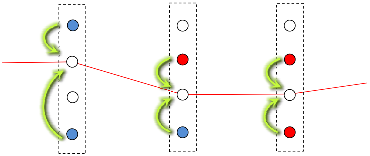
- Start with a mapping p that substitutes *everything* with y
- Auxiliary problem $g = (I - P^{\mathsf{T}})f$
- Check relaxed-improving conditions by solving LP dual:

$$\max_{\varphi} g^{\varphi}_{\varnothing} \overset{?}{\geq} 0$$

$$(\forall \mathrm{C} \in \mathcal{E}, \forall x_{\mathrm{C}})\ g^{\varphi}_{\mathrm{C}}(x_{\mathrm{C}}) \geq 0$$

- If not satisfied, determine active (blocking) constraints on nodes
- Release blocking constraints by pruning some maps

Discrete Cutting Plane



Algorithm

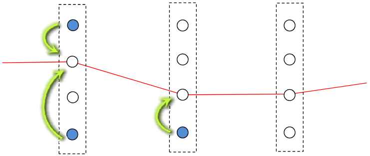
- Start with a mapping p that substitutes *everything* with y
- Auxiliary problem $g = (I - P^{\mathsf{T}})f$
- Check relaxed-improving conditions by solving LP dual:

$$\max_{\varphi} g^{\varphi}_{\varnothing} \quad \overset{?}{\geq} \quad 0$$

$$(\forall \mathrm{C} \in \mathcal{E},\ \forall x_{\mathrm{C}})\ g^{\varphi}_{\mathrm{C}}(x_{\mathrm{C}}) \geq 0$$

- If not satisfied, determine active (blocking) constraints on nodes
- Release blocking constraints by pruning some maps

Discrete Cutting Plane



Algorithm

- Start with a mapping p that substitutes *everything* with y
- Auxiliary problem $g = (I - P^{\mathsf{T}})f$
- Check relaxed-improving conditions by solving LP dual:

$$\max_{\varphi} g^{\varphi}_{\varnothing} \quad \overset{?}{\geq} \quad 0$$

$$(\forall \mathrm{C} \in \mathcal{E},\ \forall x_{\mathrm{C}})\ g^{\varphi}_{\mathrm{C}}(x_{\mathrm{C}}) \geq 0$$

- If not satisfied, determine active (blocking) constraints on nodes
- Release blocking constraints by pruning some maps

Correctness and Optimality

Properties:

- Runs in polynomial time;
- Solves the maximum persistency problem exactly when the dual is solved exactly and the solution is a strict relative interior optimal (e.g. interior point method);
- Returns an improving mapping even when the dual is solved sub-optimally – can use fast dual solvers, we used TRW-S by Kolmogorov (2006).
- Generalizes LP approach by Shekhovtsov (2014) and pruning Swoboda et al. (2014) (CVPR'14).

Efficiency

- solving relaxed inference approximately even once is slow
- TRW-S is not finitely converging

How can we iterate such relaxed inference?

Efficiency

- solving relaxed inference approximately even once is slow
- TRW-S is not finitely converging

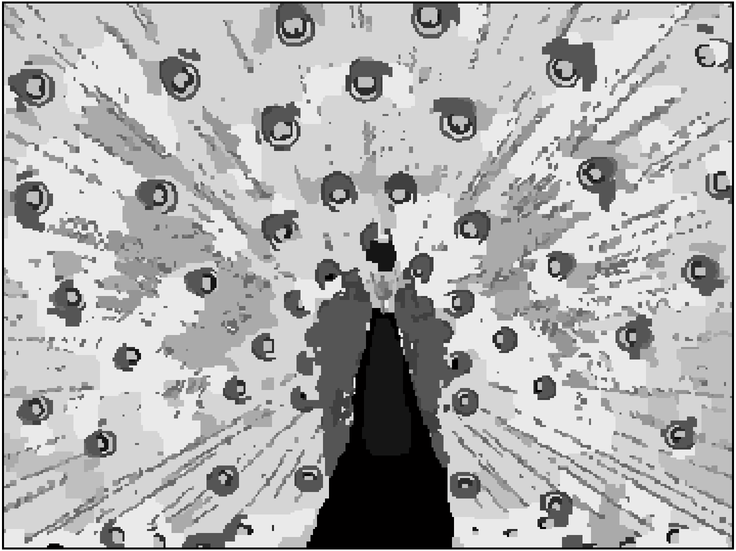
How can we iterate such relaxed inference?

Fast implementation with TRW-S

- Incremental: reuse reparametrizations φ
- Guaranteed to prune something even after 1 iteration of TRW-S (there is a blocking constraint not yet pruned)
- An optimal pruning is often possible before the dual is solved (cuts)
- Problem reductions preserving the sufficient condition
- Fast message passing for $(I - P^{\mathsf{T}})f$ with reductions

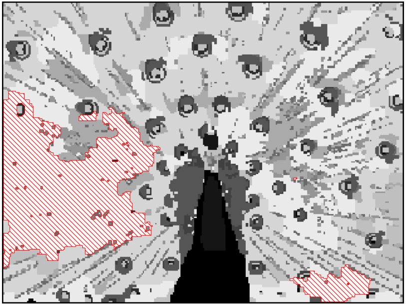
Instance	Initialization (1000 it.)	Extra time for persistency				
		no speedups	+reduction	+node pruning	+labeling pruning	+fast msgs
Protein folding `1CKK`	8.5s	268s (26.53%)	168s (26.53%)	2.0s (26.53%)	2.0s (26.53%)	2.0s (26.53%)
colorseg-n4 `pfau-small`	9.3s	439s (88.59%)	230s (93.41%)	85s (93.41%)	76s (93.41%)	19s (93.41%)

Algorithm Video

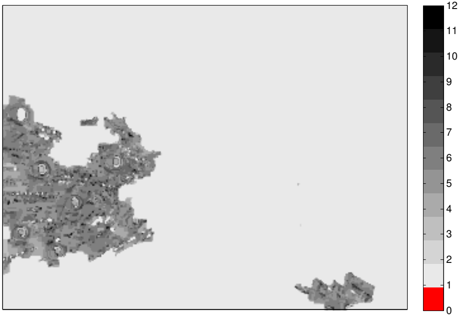


Algorithm Output

Proved optimal part



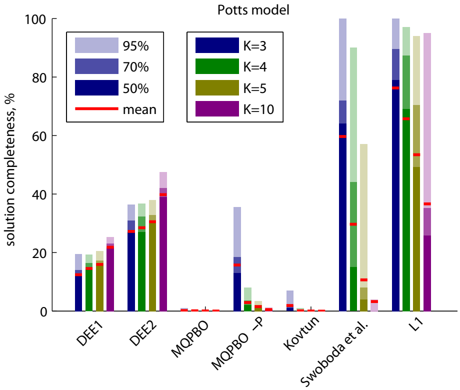
Reminder



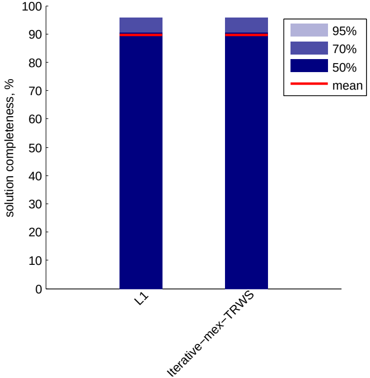
number of labels remained

Evaluation: Random Problems

CVPR'14 Comparison

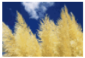


L1 vs. new algorithm using TRW-S



- Algorithm using TRW-S performs closely to maximum persistency

OpenGM Benchmark



ColorSegmentation (N8)
J. Lellmann et.al.
converted by J. Lellmann and J.H. Kappes

Object Segmentation
K. Alahari et.al.
converted by J.H. Kappes

MRF Stereo
R. Szeliski et.al.
converted by J.H. Kappes

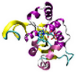
ColorSegmentation
K. Alahari et.al.
converted by J.H. Kappes

MRF Photomontage
R. Szeliski et.al.
converted by J.H. Kappes

MRF Inpainting
R. Szeliski et.al.
converted by J.H. Kappes

Chinese Characters
S. Nowozin et.al.
converted by S. Nowozin and J. H. Kappes

Scene Decomposition
Gould et.al.
converted by S. Nowozin and J. H. Kappes



Protein Folding
Yanover et. al.
converted by Joerg Kappes

Brain 3mm
J. H. Kappes et.al.
converted by J. H. Kappes

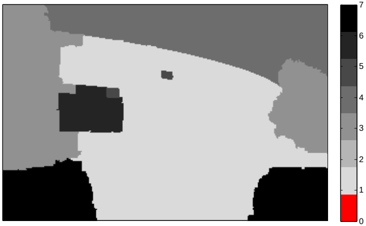
Geometric Surface Labeling (3)
Gallagher et.al.
converted by D. Batra and J. H. Kappes

Problem family	#I	#L	#V	MQPBO		MQPBO-10		Kovtun		[29]-TRWS		Our-TRWS	
mrf-stereo	3	16-60	> 100000	†		†		†		2.5h	13%	117s	**73.56%**
mrf-photomontage	2	5-7	≤ 514080	93s	22%	866s	16%	†		3.7h	16%	483s	**41.98%**
color-seg	3	3-4	≤ 424720	22s	11%	87s	16%	**0.3s**	98%	1.3h	>**99%**	61.8s	**99.95%**
color-seg-n4	9	3-12	≤ 86400	22s	8%	398s	14%	**0.2s**	67%	321s	90%	4.9s	**99.26%**
ProteinFolding	21	≤ 483	≤ 1972	685s	2%	2705s	2%	†		48s	18%	9.2s	**55.70%**
object-seg	5	4-8	68160	3.2s	0.01%	†		**0.1s**	93.86%	138s	98.19%	2.2s	**100%**

OpenGM Benchmark: Easy Examples

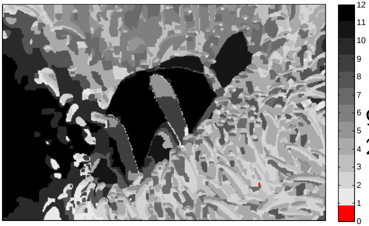
- Some problems are easy (TRWS finds optimal solution or near)

Object Segmentation



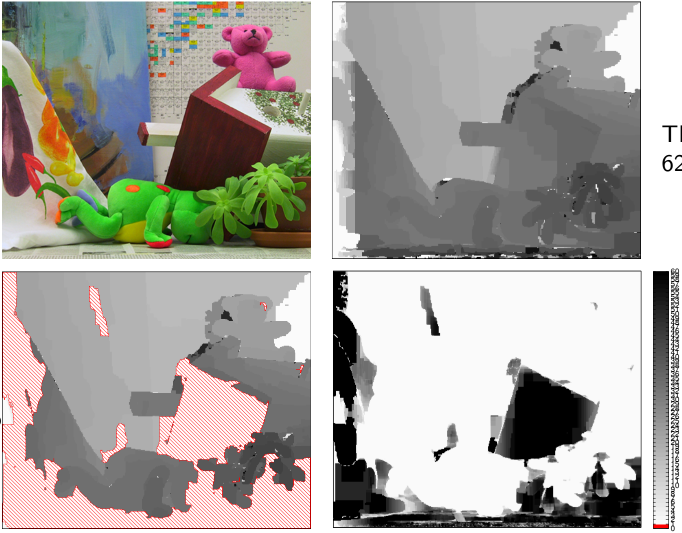
100%
1.7s

Color Segmentation



99.99%
2.8s

OpenGM Benchmark: Hard Examples

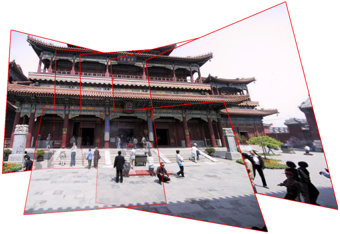


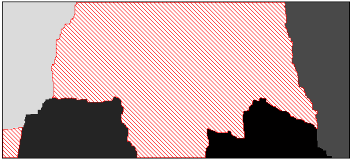
TRW-S
62s

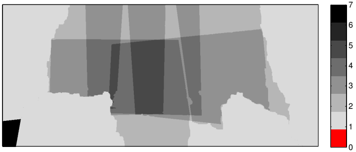
+180s
75%

OpenGM Benchmark: Very Hard Examples

Panorama Stitching







Panorama Stitching with Constraints

Conclusion

- New general sufficient condition (local + equivalent transforms)
- Covers many methods in the literature (! does not imply it is very powerful)
- Developed an efficient algorithm (implementation available, matlab interface)
- Algorithm can be understood as converting a method without guarantees (TRW-S) into a method with guarantees at a reasonable overhead
- What are these guarantees useful for further? Model verification? Learning?

Thank You!

M. L. Balinski, "Integer programming: Methods, uses, computations," *Management Science*, vol. 12, no. 3, pp. 253–313, 1965.

L.-C. Lorentzen, "Notes on covering of arcs by nodes in an undirected graph," 1966.

G. L. Nemhauser and L. E. Trotter, Jr., "Vertex packings: Structural properties and algorithms," *Mathematical Programming*, vol. 8, pp. 232–248, 1975.

J.-C. Picard and M. Queyranne, "On the integer-valued variables in the linear vertex packing problem," *Mathematical Programming*, vol. 12, no. 1, pp. 97–101, 1977. [Online]. Available: http://dx.doi.org/10.1007/BF01593772

P. Hammer, P. Hansen, and B. Simeone, "Roof duality, complementation and persistency in quadratic 0-1 optimization," *Mathematical Programming*, pp. 121–155, 1984.

E. Boros, P. L. Hammer, and X. Sun, "Network flows and minimization of quadratic pseudo-Boolean functions," RUTCOR, Tech. Rep. RRR 17-1991, May 1991.

H. D. Sherali and W. P. Adams, "A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems," *SIAM J. Discrete Math.*, vol. 3, no. 3, pp. 411–430, 1990.

V. Kolmogorov, "Generalized roof duality and bisubmodular functions," *Discrete Applied Mathematics*, vol. 160, no. 4-5, pp. 416–426, 2012.

I. Kovtun, "Partial optimal labeling search for a NP-hard subclass of (max, +) problems," in *DAGM-Symposium*, 2003, pp. 402–409.

R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *PAMI*, vol. 30, no. 6, pp. 1068–1080, 2008.

M. Shlezinger, "Syntactic analysis of two-dimensional visual signals in the presence of noise," *Cybernetics and Systems Analysis*, vol. 4, pp. 113–130, 1976, see review by Werner (2007).

M. Wainwright, T. Jaakkola, and A. Willsky, "Exact MAP estimates by (hyper)tree agreement," in *NIPS*, S. T. S. Becker and K. Obermayer, Eds. MIT Press, 2003, pp. 809–816.

A. Shekhovtsov, "Maximum persistency in energy minimization," in *CVPR*, June 2014, pp. 1162–1169.

——, "Higher order maximum persistency and comparison theorems," *Optimization Online e-prints*, 2014, under review in CVIU.

R. F. Goldstein, "Efficient rotamer elimination applied to protein side-chains and related spin glasses," *Biophysical Journal*, vol. 66, no. 5, pp. 1335–1340, May 1994.

P. Kohli, A. Shekhovtsov, C. Rother, V. Kolmogorov, and P. Torr, "On partial optimality in multi-label MRFs," in *ICML*, A. McCallum and S. Roweis, Eds. Omnipress, 2008, pp. 480–487.

I. Kovtun, "Sufficient condition for partial optimality for (max, +) labeling problems and its usage," *Control Systems and Computers*, no. 2, 2011, special issue.

P. Swoboda, B. Savchynskyy, J. Kappes, and C. Schnrr, "Partial optimality by pruning for MAP-inference with general graphical models," in *CVPR*, June 2014, pp. 1170–1177.

H. Ishikawa, "Transformation of general binary MRF minimization to the first-order case," *PAMI*, vol. 33, no. 6, pp. 1234–1249, June 2011.

A. Fix, A. Gruber, E. Boros, and R. Zabih, "A graph cut algorithm for higher-order Markov random fields," in *ICCV*, 2011, pp. 1020–1027.

V. Kolmogorov, "Generalized roof duality and bisubmodular functions," in *NIPS*, 2010, pp. 1144–1152.

F. Kahl and P. Strandmark, "Generalized roof duality for pseudo-Boolean optimization," in *ICCV*, 2011, pp. 255–262.

W. P. Adams, J. B. Lassiter, and H. D. Sherali, "Persistency in 0-1 polynomial programming," *Mathematics of Operations Research*, vol. 23, no. 2, pp. 359–389, Feb. 1998.

T. Werner, "A linear programming approach to max-sum problem: A review," *PAMI*, vol. 29, no. 7, pp. 1165–1179, July 2007.

J. Thapper and S. Živný, "The complexity of finite-valued CSPs," in *Symposium on the Theory of Computing (STOC)*. ACM, 2013, pp. 695–704. [Online]. Available: http://www.cs.ox.ac.uk/Stanislav.Zivny/homepage/publications/tz13stoc-preprint.pdf

V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *PAMI*, vol. 28(10), pp. 1568–1583, october 2006.

A. Shekhovtsov and V. Hlaváč, "On partial opimality by auxiliary submodular problems," *Control Systems and Computers*, no. 2, 2011, special issue.

A. Shekhovtsov, V. Kolmogorov, P. Kohli, V. Hlavac, C. Rother, and P. Torr, "LP-relaxation of binarized energy minimization," Czech Technical University, Tech. Rep. CTU–CMP–2007–27, 2008.

P. Swoboda, B. Savchynskyy, J. Kappes, and C. Schnörr, "Partial optimality via iterative pruning for the Potts model," in *SSVM*, 2013.

K. Alahari, P. Kohli, and P. H. S. Torr, "Reduce, reuse & recycle: Efficiently solving multi-label MRFs," in *CVPR*, 2008.

M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1-2, pp. 1–305, 2008.

A. M. Koster, S. Hoesel, and A. Kolen, "The partial constraint satisfaction problem: Facets and lifting theorems," *Operations Research Letters*, vol. 23, pp. 89–97(9), 1998.

C. Chekuri, S. Khanna, J. Naor, and L. Zosin, "Approximation algorithms for the metric labeling problem via a new linear programming formulation," in *In Symposium on Discrete Algorithms*, 2001, pp. 109–118.

C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer, "Optimizing binary MRFs via extended roof duality," in *CVPR*, 2007.

I. Kovtun, "Image segmentation based on sufficient conditions of optimality in NP-complete classes of structural labelling problem," Ph.D. dissertation, IRTC ITS National Academy of Sciences, Ukraine, 2004, in Ukrainian.

E. Boros and P. Hammer, "Pseudo-Boolean optimization," *Discrete Applied Mathematics*, vol. 1-3, no. 123, pp. 155–225, 2002.

J. Desmet, M. D. Maeyer, B. Hazes, and I. Lasters, "The dead-end elimination theorem and its use in protein side-chain positioning," *Nature*, vol. 356, pp. 539–542, 1992.

P. Orponen and H. Mannila, "On approximation preserving reductions: Complete problems and robust measures (revised version)," Tech. Rep., 1990.