Heidelberg Collaboratory for Image Processing

Getting Feasible Variable Estimates From Infeasible Ones: MRF Local Polytope Study

Bogdan Savchynskyy, Stefan Schmidt
Heidelberg Collaboratory for Image Processing, University of Heidelberg, Germany
e-mail: bogdan.savchynskyy@iwr.uni-heidelberg.de

Abstract

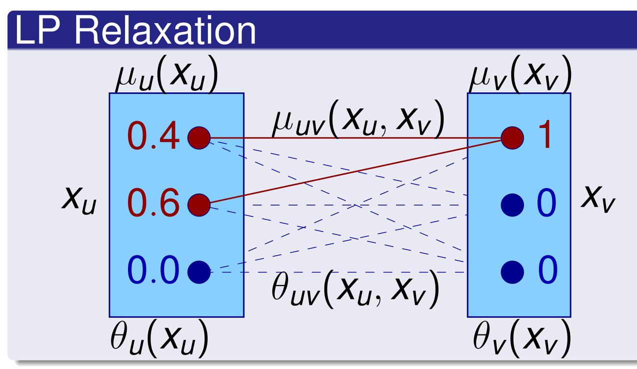
- We consider local polytope relaxation of the energy minimization/MAP-inference problem for undirected graphical models.
- **We propose** a novel method for constructing *feasible relaxed* primal estimates for a range of dual and saddle-point algorithms addressing this and similar problems with a separable structure.
- **We provide** a theoretical analysis of our method, which suggest its *optimality* with respect to all other projections.

MAP-Inference Problem

Given the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, associated variables $x_v \in \mathcal{X}_v$, $v \in \mathcal{V}$, and potentials $\theta_{w,x_w} \in \mathbb{R}$, $w \in \mathcal{V} \cup \mathcal{E}$, we consider the energy minimization problem

$$\min_{x\in\mathcal{X}} E(\theta,x) = \min_{x\in\mathcal{X}} \left\{ \sum_{v\in\mathcal{V}} \theta_{v,x_v} + \sum_{uv\in\mathcal{E}} \theta_{uv,x_{uv}} \right\} = \min_{x\in\mathcal{X}} \langle \theta, \delta(x)\rangle = \min_{\mu\in \mathrm{conv}(\delta(\mathcal{X}))} \langle\theta,\mu\rangle \geq \min_{\mu\in\mathcal{L}\supset \mathrm{conv}(\delta(\mathcal{X}))} \langle\theta,\mu\rangle$$

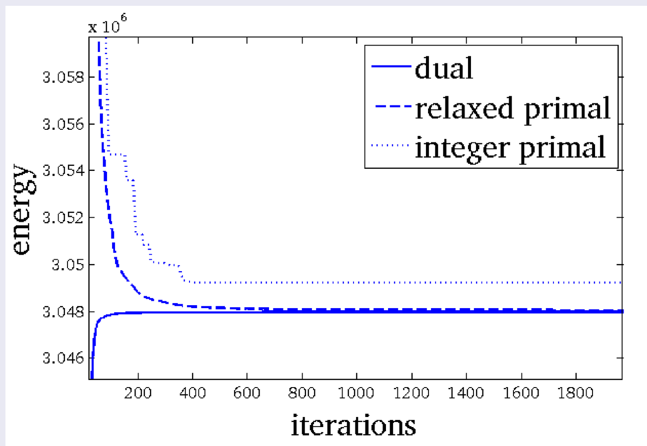
LP Relaxation



Local polytope (LP) relaxation: $\min_{\mu\in\mathcal{L}} \langle\theta,\mu\rangle$, where

$$\mathcal{L} = \left\{ \begin{array}{ll} \sum_{x_v}\mu_v = 1 & , v\in\mathcal{V} \\ \sum_{x_u}\mu_{uv,x_{uv}} = \mu_{v,x_v} & , v\in\mathcal{V},\ x_v\in\mathcal{X}_v \\ \sum_{x_v}\mu_{uv,x_{uv}} = \mu_{u,x_u} & , u\in\mathcal{V},\ x_u\in\mathcal{X}_u \\ \mu \geq 0 & \end{array} \right\} \supset \mathrm{conv}(\delta(\mathcal{X}))$$

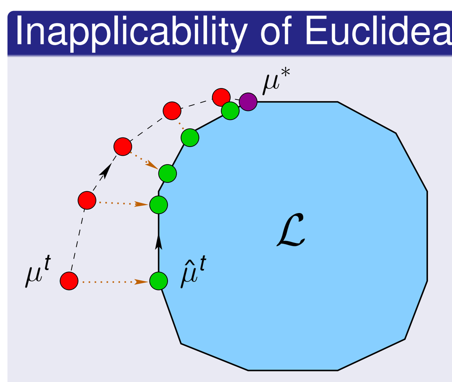
Motivation for Relaxed Feasible Primal Estimates



Duality+integrality gap between dual and integral primal

Feasible relaxed primal estimates
- provide a theoretically sound stopping condition,
- provide a basis for the comparison of different optimization schemes,
- enable *adaptive* optimization schemes depending on the duality gap, e.g. step-size selection in subgradient algorithm,
- enable tightening of relaxations with primal cutting-plane based approaches.

Inapplicability of Euclidean Projection for Feasibility of Primal Relaxed Estimates

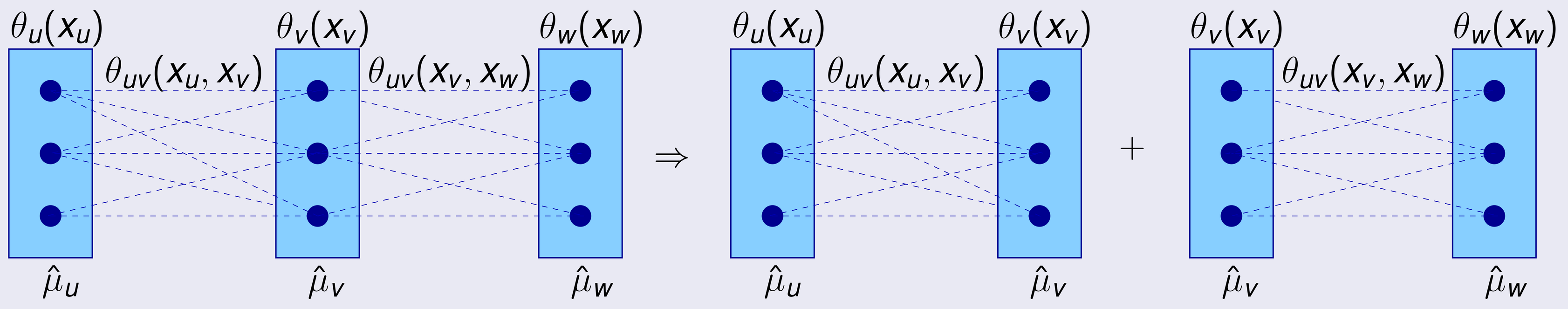


- Many algorithms produce infeasible primal estimates $\mu^t \xrightarrow{t\to\infty} \mu^*$, $\mu^t \notin \mathcal{L}$
- Infeasible estimates μ^t are almost useless.
- Their (Euclidean) projection to $\mathcal{L}$, $\min_{\mu\in\mathcal{L}} \|\mu^t - \mu\|^2$, is computationally expensive!

Separability of the Local Polytope w.r.t. Fixed μ_v, $v \in \mathcal{V}$

Let $\mu_v = \hat{\mu}_v, v \in \mathcal{V}$, then $\mathcal{L}_{uv}(\hat{\mu}_u,\hat{\mu}_v) := \{\mu_{uv}\geq 0\colon \sum_{x_u}\mu_{uv,x_{uv}} = \hat{\mu}_{v,x_v},\ \sum_{x_v}\mu_{uv,x_{uv}} = \hat{\mu}_{u,x_u}\}$

$$\arg\min_{\substack{\mu\in\mathcal{L} \\ \mu_v=\hat{\mu}_v, v\in\mathcal{V}}} \langle\theta,\mu\rangle = \left(\arg\min_{\mu_{uv}\in\mathcal{L}_{uv}(\hat{\mu}_u,\hat{\mu}_v)} \langle\theta_{uv},\mu_{uv}\rangle \quad , \quad \arg\min_{\mu_{vw}\in\mathcal{L}_{vw}(\hat{\mu}_v,\hat{\mu}_w)} \langle\theta_{vw},\mu_{vw}\rangle \right)$$



Optimizing Projection:
$$\left\{ \begin{array}{lll} \hat{\mu}_v = & \Pi_{\Delta_{|\mathcal{X}_v|}}(\mu_v) & , v\in\mathcal{V} \quad \text{\# feasibility, in simple case } \hat{\mu}_v := \mu_v \\ \hat{\mu}_{uv} = & \arg\min_{\mu_{uv}\in\mathcal{L}_{uv}(\hat{\mu}_u,\hat{\mu}_v)} \langle\theta_{uv},\mu_{uv}\rangle & , uv\in\mathcal{E} \quad \text{\# optimization w.r.t. } \mu_{uv},\ uv\in\mathcal{E} \end{array} \right.$$

Theorem. From $E(\mu^t) \to E^*_{\mathcal{L}}$ and $\mu^t \to \mathcal{L}$ follows $E(\hat{\mu}^t) \to E^*_{\mathcal{L}}$. Moreover:

$$|E(\hat{\mu}) - E^*_{\mathcal{L}}| \leq |E(\mu) - E^*_{\mathcal{L}}| + (\|\theta_{\mathcal{V}}\| + \|\theta_{\mathcal{E}}\|)\|\mu - \Pi_{\mathcal{L}}(\mu)\|$$

$\Rightarrow$ Projected estimates $\hat{\mu}^t$ converge with the same rate as the non-projected ones μ^t.

Efficiently Computing the Optimizing Projection for the Local Polytope

$$\min_{\mu_{uv}\geq 0} \langle\theta_{uv},\mu_{uv}\rangle \quad \text{s.t.} \quad \sum_{x_u}\mu_{uv,x_{uv}} = \hat{\mu}_{v,x_v},\ \sum_{x_v}\mu_{uv,x_{uv}} = \hat{\mu}_{u,x_u}$$

- Small-sized *standard, well-studied* transportation problem
- 0.01-0.02 sec for in total 256×256 problems, 4×4 labels each
- code is publicly available in OpenGM library [1]

Beyond the Local Polytope: Optimizing Projection for Arbitrary Convex Functions and Sets

$f^*_C = \arg\min_{z\in C} f(z)$ - convex; $C \subset \mathbb{R}^n$ - convex closed; $z \equiv (x, y)$

Optimizing projection
$$\left\{ \begin{array}{ll} \hat{x} := \Pi_{C_X}(x),\ C_X = \{x \mid \exists y\colon (x,y)\in C\} & \text{\# feasibility, in simple case } \hat{x} := x \\ \hat{y} := \arg\min_{y\colon(\hat{x},y)\in C} f(\hat{x},y) & \text{\# optimization w.r.t. } y \text{ for a fixed } x \end{array} \right.$$

Theorem

$f(\hat{x}^t,\hat{y}^t) \xrightarrow{t\to\infty} f^*_C$, f-continuous and $f(z^t)\to f^*_C$, $z^t \to C$

$|f(\hat{x},\hat{y}) - f^*| \leq |f(x,y) - f^*| + (L_X(f) + L_Y(f))\|z - \Pi_C(z)\|$, f-Lipschitz-continuous w.r.t. x and y

$|f(\hat{x},\hat{y}) - f^*| \leq |f(x,y) - f^*| + L_Y(f)\|z - \Pi_C(z)\|$, f-Lipschitz-continuous w.r.t. y and $\hat{x} = x$.

Comparable to the Euclidean projection: $|f(\hat{x},\hat{y}) - f^*| \leq |f(x,y) - f^*| + L_{XY}(f)\|z - \Pi_D(z)\|$

Optimality of Optimizing Projection

Theorem $f(\hat{x},\hat{y}) \leq f(x,y)$ for any $(x,y)\in C$ and $f(\hat{x},\hat{y}) < f(x,y)$ if $y \notin \arg\min_{y'\colon(x,y')\in C} f(x,y')$. # Optimizing projection improves ANY other projection.

Optimizing Projection for a Smoothed Primal (Approximation of the Partition Function)

$\min_{\mu\in\mathcal{L}} A(\mu) := \langle\theta,\mu\rangle - \sum_{w\in\mathcal{V}\cup\mathcal{E}} b_w H(\mu_w)$ - partition function approximation [4, 3]

Let
$$\left\{ \begin{array}{lll} \hat{\mu}_v = & \Pi_{\Delta_{|\mathcal{X}_v|}}(\mu_v) & , v\in\mathcal{V} \quad \text{\# feasibility of unary marginals} \\ \hat{\mu}_{uv} = & \arg\min_{\mu_{uv}\in\mathcal{L}_{uv}(\hat{\mu}_u,\hat{\mu}_v)} \langle\theta_{uv} - b_{uv}\log(\mu_{uv}),\mu_{uv}\rangle & , uv\in\mathcal{E} \quad \text{\# optimization w.r.t. p/w marginals} \end{array} \right.$$

Corollary From $A(\mu^t) \xrightarrow{t\to\infty} A^*_{\mathcal{L}}$, $\mu^t \xrightarrow{t\to\infty} \mathcal{L}$ follows $A(\hat{\mu}^t) \xrightarrow{t\to\infty} A^*_{\mathcal{L}}$.

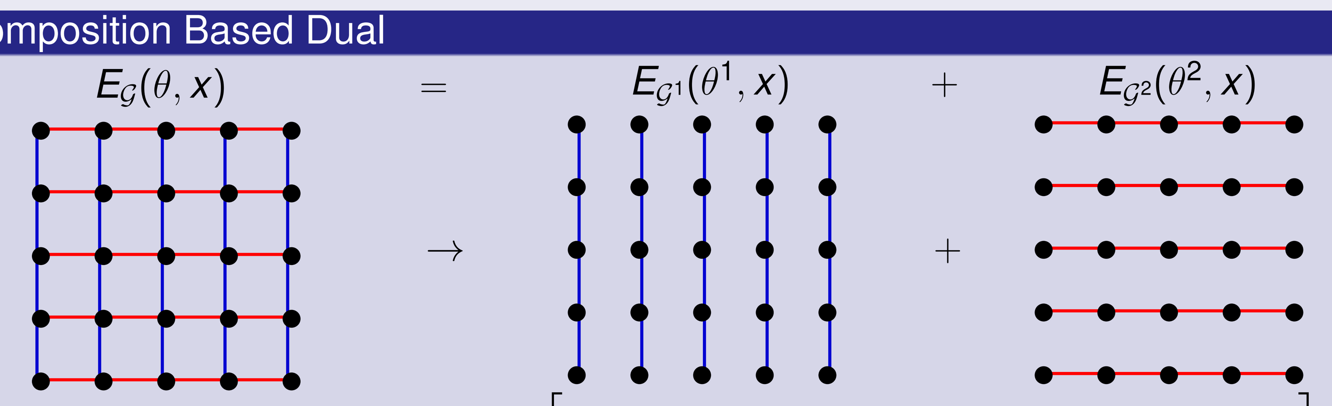
Getting Infeasible Primal Estimates

Primal-Dual Algorithms (incl. ADMM, ADLP): [2, 11, 7, 8]

$$\max_{\mu\geq 0}\min_{\phi} \quad \langle -b,\phi\rangle + \left\langle \mu, A^\top\phi \right\rangle - \langle\theta,\mu\rangle + \tau\|A\mu - b\|^2.$$

$\mu^t \notin \mathcal{L}$ - *infeasible* primal iterates of the algorithms. Projected iterates $\hat{\mu}^t \in \mathcal{L}$ are *feasible*.

Dual Decomposition Based Dual



$$\min_{x\in\mathcal{X}} E_{\mathcal{G}}(\theta,x) \geq \max_{\theta^1+\theta^2=\theta} \left[\min_{x^1\in\mathcal{X}} E_{\mathcal{G}^1}(\theta^1,x^1) + \min_{x^2\in\mathcal{X}} E_{\mathcal{G}^2}(\theta^2,x^2) \right]$$

Parametrization: $\theta^1_{\mathcal{V}}(\lambda) = \frac{\theta_{\mathcal{V}}}{2} + \lambda_{\mathcal{V}}$, $\theta^2_{\mathcal{V}}(\lambda) = \frac{\theta_{\mathcal{V}}}{2} - \lambda_{\mathcal{V}}$, $\theta^i_{uv}(\lambda) = \theta_{uv}$, $uv \in \mathcal{E}^i$

Dual: $U(\lambda) = \sum_{i=1}^2 \min_{x^i\in\mathcal{X}} E_{\mathcal{G}^i}(\theta^i(\lambda), x^i)$ # piecewise linear, concave, non-smooth

$\forall\mu\in\mathcal{L}, \lambda\in\mathbb{R}^n\ E(\mu) \geq U(\lambda)$. In optimum: $(\lambda^*,\mu^*) \Rightarrow E(\mu^*) = U(\lambda^*)$.

Subgradient Based Algorithms [6, 5]

Subgradient: $\frac{\partial U}{\partial\lambda} = \delta_{\mathcal{V}}(x^{*1}) - \delta_{\mathcal{V}}(x^{*2})$

Algorithm: $\lambda^{t+1} = \lambda^t + \tau^t\frac{\partial U}{\partial\lambda}(\lambda^t)$ $\quad \tau^t \to 0,\ \sum_{t=1}^{\infty}\tau^t = \infty$

Primal sequences: $i = 1,2:$ $\frac{\sum_{k=1}^t \delta_{\mathcal{V}}(x^{*i,k})}{t} \xrightarrow{t\to\infty} \mu^*_{\mathcal{V}}$ # time-averaged subgradient

$i = 1,2:$ $\frac{\sum_{k=1}^t \tau^k\delta_{\mathcal{V}}(x^{*i,k})}{\sum_{k=1}^t\tau^k} \xrightarrow{t\to\infty} \mu^*_{\mathcal{V}}$ # step-size-averaged subgradient

Bundle method [5]: $i = 1,2:$ $\frac{\sum_{k=1}^t \xi^k\delta_{\mathcal{V}}(x^{*i,k})}{\sum_{k=1}^t\xi^k} \xrightarrow{t\to\infty} \mu^*_{\mathcal{V}}$ # bundle-averaged subgradient

Smoothing Based Algorithms [9, 10, 3]

Smoothed dual: $\tilde{U}_\rho(\lambda) + 2\rho\log|\mathcal{X}| \geq U(\lambda) \geq \tilde{U}_\rho(\lambda)$

Unary marginals: $i = 1,2$ $\quad \mu^i_\rho(\lambda)_{v,x_v} := \frac{\sum_{x'\in\mathcal{X}, x'_v=x_v}\exp\left\langle -\theta^i(\lambda)/\rho, \phi(x')\right\rangle}{\exp(-\tilde{U}^i_\rho(\lambda)/\rho)}$

Gradient: $\nabla\tilde{U}_\rho(\lambda) = \mu^1_\rho(\lambda)_{\mathcal{V}} - \mu^2_\rho(\lambda)_{\mathcal{V}}$

Primal sequences: $i = 1,2:$ $\mu^i_\rho(\lambda^t)_v \xrightarrow{\lambda^t\to\hat{\lambda}^*} \tilde{\mu}^*_v = \arg\min_{\mu\in\mathcal{L}} A(\theta,\mu)$ # gradient

Experimental Evaluation

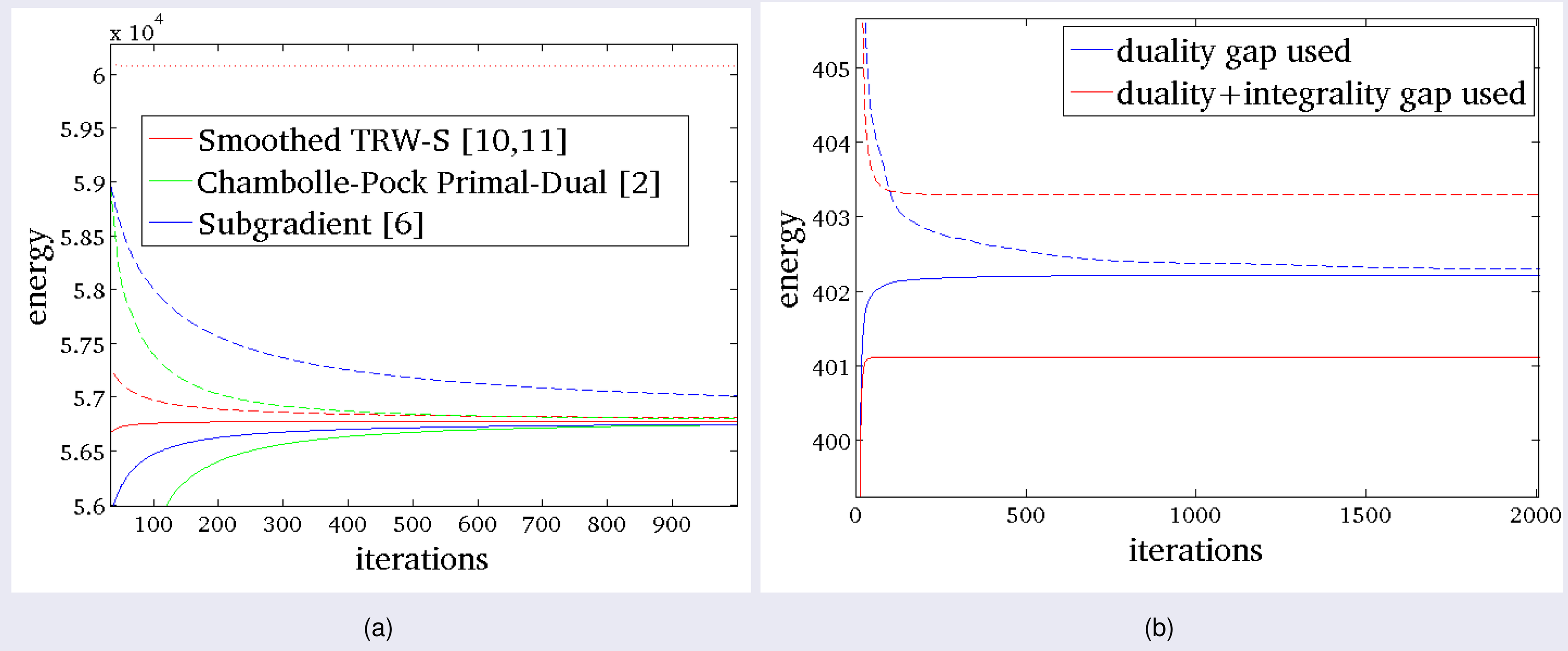


Figure : **(a)** Convergence of the primal (dashed lines) and dual (solid lines) bounds to the same optimal limit value for primal-dual Chambolle-Pock [2, 11], smoothing-based Smoothed TRW-S [10], and Subgradient [6] algorithms. The obtained integer bound is plotted as a dotted line. Randomly generated 256×256, 4 labels grid model. **(b)** Adaptively updating smoothing based on duality gap and duality+integrality gap in the Smoothed TRW-S [10] algorithm. Randomly generated 25×25, 4 labels grid model.

References

[1] Bjoern Andres, Thorsten Beier, and Joerg H. Kappes. OpenGM: A C++ library for discrete graphical models. Technical report, arXiv:1206.0111, 2012.

[2] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, pages 1–26, 2010.

[3] T. Hazan and A. Shashua. Norm-product belief propagation: Primal-dual message-passing for approximate inference. *IEEE Trans. on Inf. Theory,*, 56(12):6294 –6316, Dec. 2010.

[4] Tom Heskes. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 16(11):2379–2413, 2004.

[5] Jörg Hendrik Kappes, Bogdan Savchynskyy, and Christoph Schnörr. A bundle approach to efficient MAP-inference by lagrangian relaxation. In *CVPR 2012*, 2012. in press.

[6] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF energy minimization and beyond via dual decomposition. *IEEE Trans. PAMI*, 33:531–552, March 2011.

[7] A. F. T. Martins, M. A. T. Figueiredo, P. M. Q. Aguiar, N. A. Smith, and E. P. Xing. An augmented Lagrangian approach to constrained MAP inference. In *ICML*, 2011.

[8] Ofer Meshi and Amir Globerson. An alternating direction method for dual MAP LP relaxation. In *ECML/PKDD (2)*, pages 470–483, 2011.

[9] Bogdan Savchynskyy, Jörg Kappes, Stefan Schmidt, and Christoph Schnörr. A study of Nesterov's scheme for Lagrangian decomposition and MAP labeling. In *CVPR 2011*, 2011.

[10] Bogdan Savchynskyy, Stefan Schmidt, Jörg Kappes, and Christoph Schnörr. Efficient MRF energy minimization via adaptive diminishing smoothing. In *UAI 2012*, pages 746–755, 2012.

[11] Stefan Schmidt, Bogdan Savchynskyy, Jörg Kappes, and Christoph Schnörr. Evaluation of a first-order primal-dual algorithm for MRF energy minimization. In *EMMCVPR 2011*, 2011.

Acknowledgement. This work has been supported by the German Research Foundation (DFG) within the program "Spatio-/Temporal Graphical Models and Applications in Image Analysis", grant GRK 1653.