A Study of Nesterov's Scheme for Lagrangian Decomposition and MAP Labeling

Bogdan Savchynskyy, Jörg Kappes, Stefan Schmidt, Christoph Schnörr

Heidelberg Collaboratory for Image Processing (HCI)
University of Heidelberg

MRF/MAP Inference – Applications

$$y^* = \arg\min_{y \in \mathcal{Y}^{\mathcal{V}}} \left[\sum_{v \in \mathcal{V}} \theta_v(y_v) + \sum_{vv' \in \mathcal{E}} \theta_{vv'}(y_v, y_{v'}) \right]$$

- **Segmentation** [Rother et al. 2004], [Nowozin, Lampert 2010]
- **Multi-camera stereo** [Kolmogorov, Zabih 2002]
- **Stereo and Motion** [Kim et al. 2003]
- **Clustering** [Zabih, Kolmogorov. 2004]
- **Medical imaging** [Raj et al. 2007]
- **Pose Estimation** [Bergtholdt et al. 2010], [Bray et al. 2006]
- ...

A comparative study of energy minimization methods *for Markov random fields with smoothness-based priors.* R. Szeliski et al. 2008

MRF/MAP Inference – Approaches

Graph Cuts

- [Boykov et al. 2001]
- [Kolmogorov, Zabih 2002]
- [Boykov, Kolmogorov 2004]

Special type of potentials. Sub-modularity

QPBO and Roof Duality

- [Hammel et al.1984],
- [Boros, Hammer 2002],
- [Rother et al. 2007], [Kohli et al. 2008]

Partial optimality.

Combinatorial methods

- [Bergtholdt et al. 2006], [Schlesinger 2009]
- [Sanchez et al. 2008],
- [Marinescu, Dechter 2009]

Exponential complexity in the worst-case.

MRF/MAP Inference – Approaches

Message passing and belief propagation

- [Weiss, Freeman 2001], [Wainwright et al. 2002], [Kolmogorov 2005], [Globerson, Jaakkola 2007]

Relaxation, dual decomposition.
Sub-optimal fixed point
Stopping criterion?

Sub-gradient Optimization Schemes

- [Komodakis et al. 2007], [Schlesinger, Giginyak 2007], [Kappes et al. 2010]

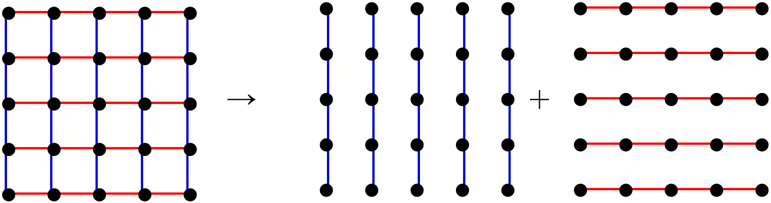
Relaxation, dual decomposition.
Slow convergence.
Stopping criterion?

Focus and Contribution:

- Local Polytope/LP relaxation based on dual decomposition – similar to message passing and sub-gradient schemes;
- efficient iterations – outperforms subgradient;
- convergence to the optimum – outperforms message passing;
- stopping criterion based on duality gap – novel!

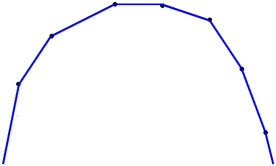
Dual Decomposition Approach

$$E(\theta, y) = E^1(\theta^1, y) + E^2(\theta^2, y)$$



$$\min_{y \in \mathcal{Y}^{\mathcal{V}}} E(\theta, y) \geqslant \max_{\theta^1 + \theta^2 = \theta} \left[\min_{y \in \mathcal{Y}^{\mathcal{V}}} E_1(\theta^1, y) + \min_{y \in \mathcal{Y}^{\mathcal{V}}} E_2(\theta^2, y) \right]$$

- Simple subproblems in parallel
- Concave, but non-smooth



Large Scale Convex Optimization

Problem: Dual Decomposition → Convex, Large-Scale, Non-Smooth

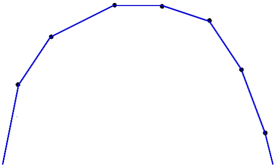
- Sub-gradient schemes: [Komodakis et al. 2007], [Schlesinger, Giginyak 2007]
- Block-coordinate ascent: [Wainwright 2004], [Kolmogorov 2005], [Globerson, Jaakkola 2007]
- Smoothing + Block-coordinate ascent: [Johnson et al. 2007], [Werner 2009]
- Proximal methods: [Ravikumar et al. 2010]
- **Smoothing technique + accelerated gradient methods: [Nesterov 2004, 2007]**
- Proximal methods: [Combettes, Wajs 2005], [Beck, Teboulle 2009]
- Proximal Primal-Dual Algorithms: [Esser et al. 2010]

Solution direction: Smooth and Optimize

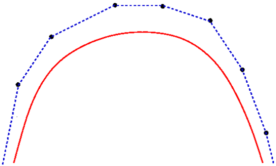
Smoothing Technique by Y.Nesterov

$$\underbrace{\min_{y \in \mathcal{D}} \left[\langle Ax, y\rangle + \phi(y)\right]}_{f(x)} \; - - - - - - \rightarrow \; \underbrace{\min_{y \in \mathcal{D}} \left[\langle Ax, y\rangle + \phi(y) + \rho d(y)\right]}_{\tilde{f}_\rho(x)}$$

Concave, but non-smooth,
convergence $t \approx O(\frac{1}{\varepsilon^2})$



Lipschitz-continuous gradient,
convergence $t \approx O(\frac{1}{\varepsilon})$

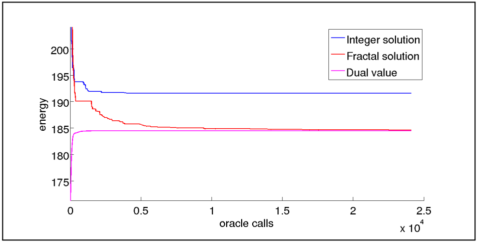


Efficient Implementation of a Nesterov's Method

	Basic scheme	**Our approach**
Stopping condition	Worst-case number of steps	Duality gap
Smoothing selection	Worst-case analysis	Adaptive
Lipschitz constant estimation (step-size selection)	Worst-case analysis	Adaptive

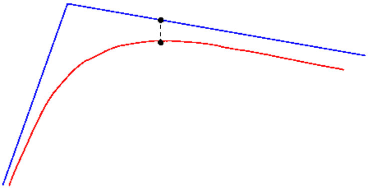
Duality Gap and Stopping Condition

$$\min_x \max_y g(x,y) - \max_y \min_x g(x,y) \leqslant \varepsilon$$

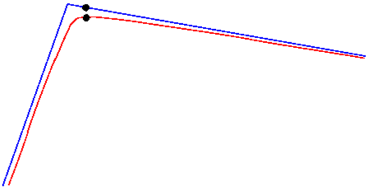


- dual decomposition approaches optimize the relaxed dual $\max_y \min_x g(x,y)$.
- standard approach – estimate a non-relaxed primal, integer solution.
- we estimate the relaxed primal $\min_x \max_y g(x,y)$ – **difficult!**.

Smoothing Selection

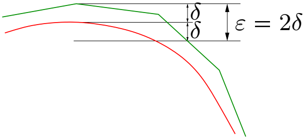


Fast optimization – low precision

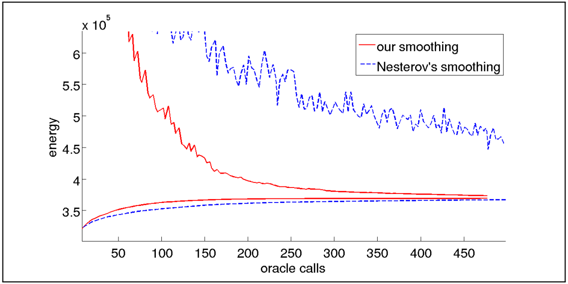


Slow optimization – high precision

Smoothing Selection



- Nesterov: worst-case estimate.
- Ours: adaptive estimate.

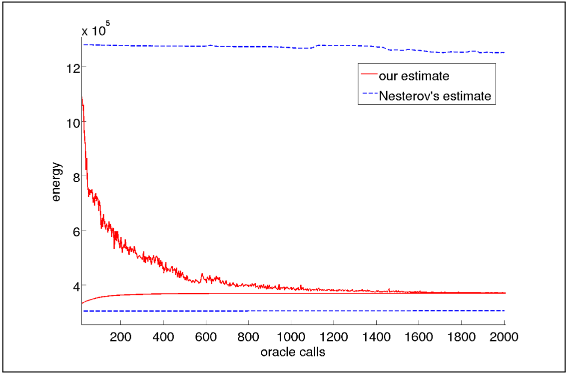


Tsukuba dataset and precision about 0.3%

Lipschitz Constant (Steps-Size) Estimation

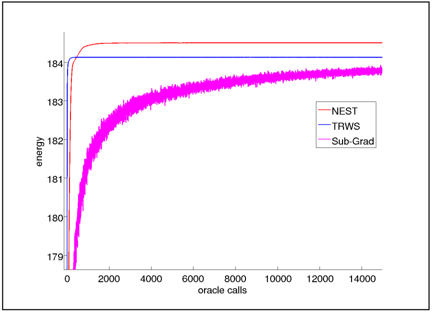
$$x = y + \frac{1}{L}\nabla f(y)$$

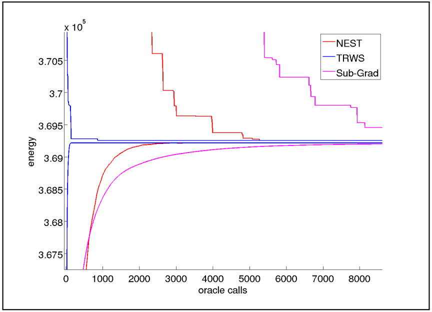
- Nesterov: worst-case estimate of L.
- Ours: adaptive estimate of L without violating the theory!



Tsukuba dataset and precision about 3%

Comparison to Other Approaches





Random **synthetic** grid model 20x20, 5 labels and **Tsukuba** dataset

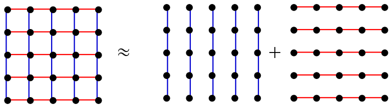
Summary

Contribution:

Improved convergence estimation: $O(\frac{1}{\varepsilon})$ vs. $O(\frac{1}{\varepsilon^2})$

Sound stopping condition: $\min_x \max_y g(x,y) - \max_y \min_x g(x,y) \leqslant \varepsilon$

Fine-grained
parallelization properties

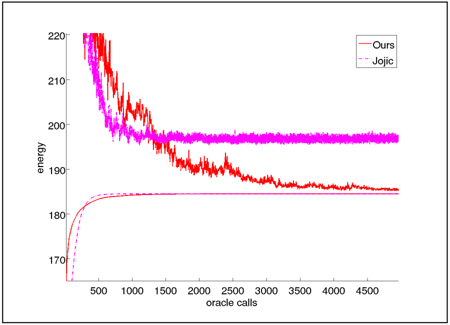


Applicable to arbitrary graphs and arbitrary potentials.

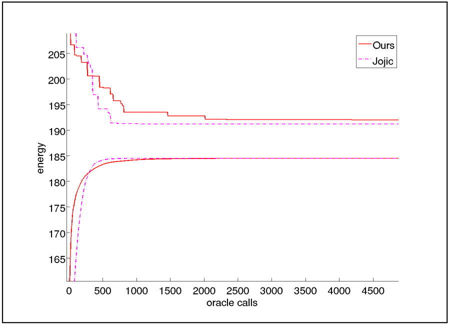
Future work:

- Examine Primal-Dual viewpoint – EMMCVPR 2011
- Appication in structured prediction and learning.

V. Jojic, S. Gould, and D. Koller. *Accelerated dual decomposition ...* 2010



Primal LP solution



Primal integer solution

Synthetic grid 20x20, 5 labels.