MPLP++: Fast, Parallel Dual Block-Coordinate Ascent (BCA) for Dense Graphical Models

Siddharth Tourani[1] Alexander Shekhovtsov [2] Carsten Rother[1] Bogdan Savchynskyy[1]

Visual Learning Lab, IWR, University of Heidelberg, Germany Centre for Machine Perception, Czech Technical University, Czech Republic

Objective

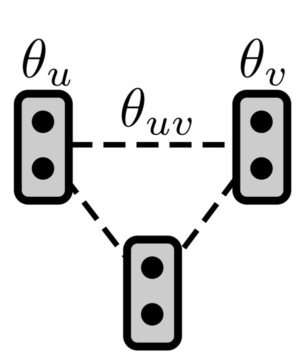
What:

- Fast and parallel algorithms for dense graphical models

Why:

- Dense graphical models are more expressive
- CNN+CRF training
- Huge datasets and Real Time Applications

MAP Inference



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

- $\theta_v \to$ node potential.
- $\theta_{uv} \to$ pairwise potential.
- $y^* \to$ optimal labelling.

Dual LP

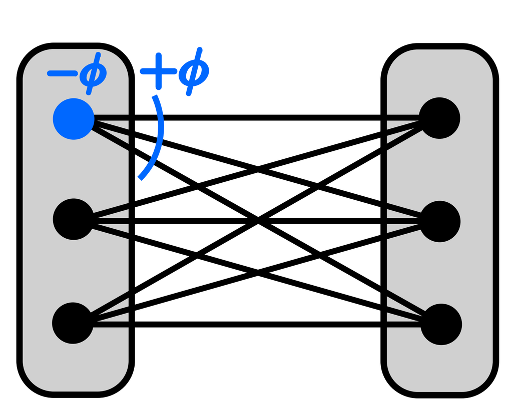
To be able to deal with arbitrary potentials, we address the dual problem:

$$D(\phi) := \sum_{u \in \mathcal{V}} \min_{s \in \mathcal{Y}} \theta^{\phi}_u(s) + \sum_{uv \in \mathcal{E}} \min_{(s,t) \in \mathcal{Y}^2} \theta^{\phi}_{uv}(s,t)\,. \qquad (1)$$

$$\theta^{\phi}_u(s) := \theta_u(s) + \textstyle\sum_{v \in \mathrm{Nb}(u)} \phi_{v \to u}(s)$$

$$\theta^{\phi}_{uv}(s,t) := \theta_{uv}(s,t) - \phi_{v \to u}(s) - \phi_{u \to v}(t)\,.$$

Dual variables $\phi_{u \to v}$ and $\phi_{v \to u}$ are the Lagrange multipliers.



$D(\phi)$ is

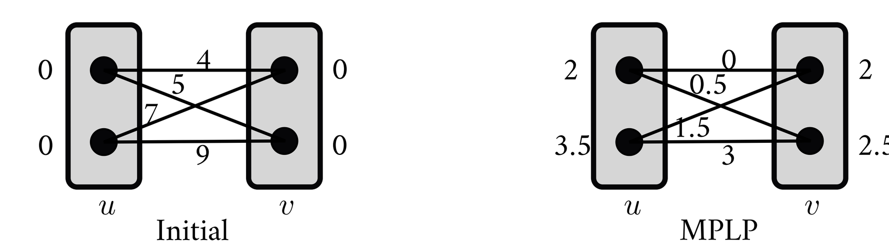
- concave
- piece-wise linear
- non-smooth

Motivation for Algorithm Structure

- The subgradient for the dual is sparse. Block Coordinate Ascent methods are more efficient than sub-gradient based methods.
- TRWS [1] is the best performer for MAP inference.
- Dense graphs need no large sub-problem decompositions.

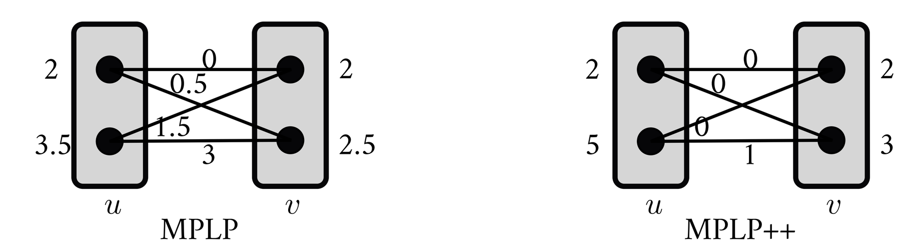
Comparing BCA Updates

$$g_{uv}(s,t) = \theta_{uv}(s,t) + \theta_u(s) + \theta_v(t)$$



MPLP: [2]

$$\begin{aligned} \theta^{\mathcal{M}}_u(s) &:= \tfrac{1}{2}\min_{t \in \mathcal{Y}} g_{uv}(s,t),\ \forall s \in \mathcal{Y}, \\ \theta^{\mathcal{M}}_v(t) &:= \tfrac{1}{2}\min_{s \in \mathcal{Y}} g_{uv}(s,t),\ \forall t \in \mathcal{Y}. \end{aligned} \qquad (\mathcal{M})$$



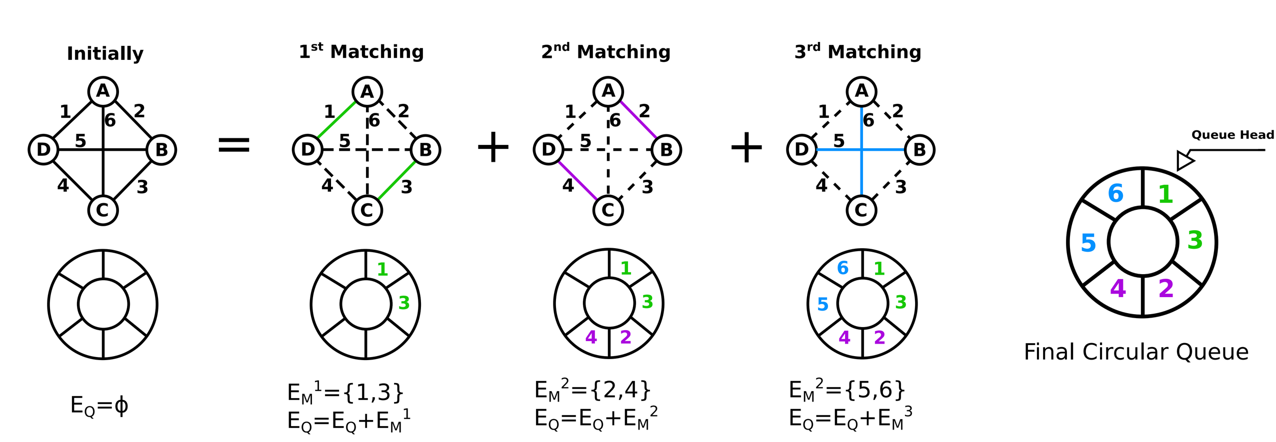
MPLP++:

$$\begin{aligned} &\theta^{\mathcal{H}}_u(s) := \theta^{\mathcal{M}}_u(s), \quad \theta^{\mathcal{H}}_v(s) := \theta^{\mathcal{M}}_v(s),\ \forall s \in \mathcal{Y}\,, \\ &\theta^{\mathcal{H}}_v(t) := \theta^{\mathcal{H}}_v(t) + \min_{s \in \mathcal{Y}}[g_{uv}(s,t) - \theta^{\mathcal{H}}_v(t) - \theta^{\mathcal{H}}_u(s)],\ \forall t \in \mathcal{Y}\,, \\ &\theta^{\mathcal{H}}_u(s) := \theta^{\mathcal{H}}_u(s) + \min_{t \in \mathcal{Y}}[g_{uv}(s,t) - \theta^{\mathcal{H}}_v(t) - \theta^{\mathcal{H}}_u(s)],\ \forall s \in \mathcal{Y}\,. \end{aligned} \qquad (\mathcal{H})$$

We prove

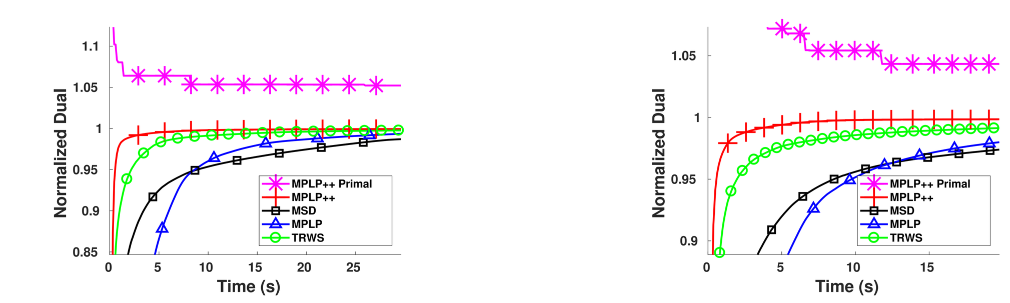
With same input, at the end of the first iteration, $D(\phi^{\mathcal{M}}) \le D(\phi^{\mathcal{H}})$.

Parallelization: CPU & GPU



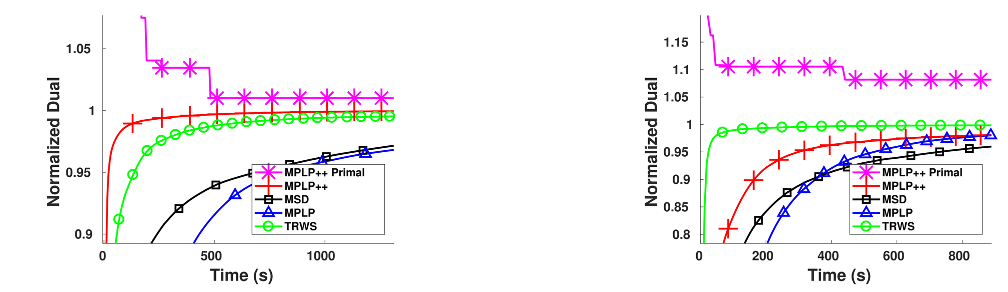
Non-incident edges can be processed in parallel. $\to$ We use maximum matching solvers.

Results



(e) protein (100% density, 40 vars, up to 200 states)

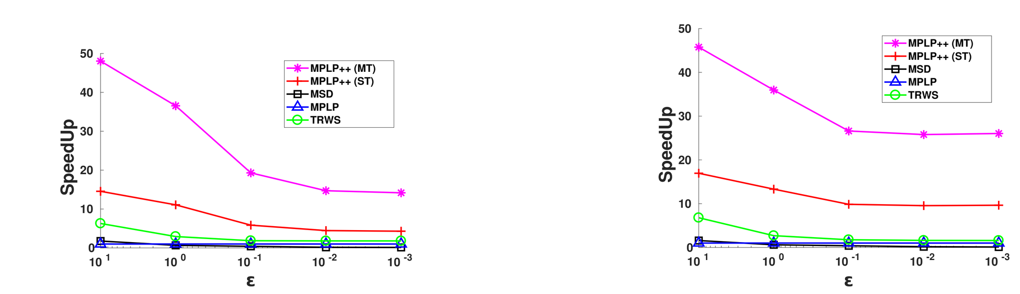
(f) worms (10% density, 556 vars, 40-50 states)



(g) pose (100% density, 600-4800 nodes,13 states)

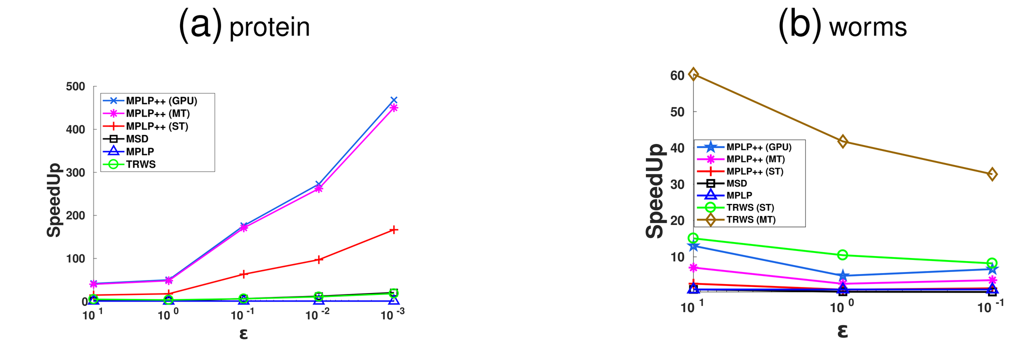
(h) stereo (tsukuba, venus, teddy, grid graph, 12-60 states)

Figure 1: Dual versus time



(a) protein

(b) worms



(c) pose

(d) stereo

Figure 2: Obtained Speedups

Conclusions and Outlook

- Our approach is the state-of-the-art method for dense graphical models ($>10\%$ graph density), beating even TRWS.
- We give CPU and GPU parallel implementations provided at **XXXX**.

References

[1] Kolmogorov, V., "Convergent tree-reweighted message passing for energy minimization", PAMI, 2006.

[2] Globerson et al., "Fixing Max-Product: Convergent Message Passing Algorithms for MAP LP-Relaxations", NIPS, 2008.

[3] A. Shekhovtsov, C. Reinbacher, G.Graber and T. Pock: Solving Dense Image Matching in Real-Time Using Discrete-Continuous Optimization, CVWW 2016.

Acknowledgements

This project has received funding from the ERC under the European Unions Horizon 2020 research and innovation program (grant agreement No 647769). A. Shekhovtsov was supported by Czech Science Foundation grant 18-25383S.