LPQP for MAP: Putting LP Solvers to Better Use

Patrick Pletscher, Sharon Wulff

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Energy Minimization and MAP Inference

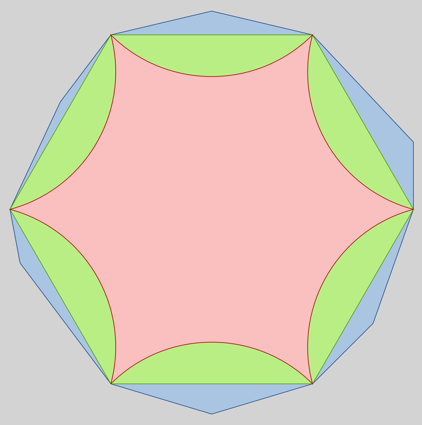
For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the pairwise MAP problem is:

$$\min_{\mathbf{x}} \sum_{i \in \mathcal{V}} \theta_i(x_i) + \sum_{(i,j)\in\mathcal{E}} \theta_{ij}(x_i, x_j) \qquad x_i \in \{1, \dots, K\}$$

- MAP for general graphs and energies is NP-hard.
- The potentials $\theta_i(x_i)$ and $\theta_{ij}(x_i, x_j)$ encode costs of an assignment.

Relaxation Approaches

MAP as an integer quadratic program:



$$\min_{\boldsymbol{\mu}} \sum_{i\in\mathcal{V}} \boldsymbol{\theta}_i^\top \boldsymbol{\mu}_i + \sum_{(i,j)\in\mathcal{E}} \boldsymbol{\mu}_i^\top \boldsymbol{\Theta}_{ij} \boldsymbol{\mu}_j$$

$$\text{s.t.} \quad \mu_{i;k} \in \{0,1\}\ \forall i,k \quad \text{and} \quad \sum_k \mu_{i;k} = 1\ \forall i.$$

Linear program (LP) relaxation:

$$\min_{\boldsymbol{\mu}\in\mathcal{L}_\mathcal{G}} \sum_{i\in\mathcal{V}} \boldsymbol{\theta}_i^\top \boldsymbol{\mu}_i + \sum_{(i,j)\in\mathcal{E}} \boldsymbol{\theta}_{ij}^\top \boldsymbol{\mu}_{ij},$$

$$\mathcal{L}_\mathcal{G} = \left\{ \boldsymbol{\mu} \,\middle|\, \begin{array}{l} \sum_k \mu_{i;k} = 1 \\ \sum_l \mu_{ij;kl} = \mu_{i;k} \\ \mu_{i;k} \geq 0\,,\ \mu_{ij;kl} \geq 0 \end{array} \right\}$$

Quadratic program (QP) relaxation:

$$\min_{\boldsymbol{\mu}\in\mathcal{L}_\mathcal{G}} \sum_{i\in\mathcal{V}} \boldsymbol{\theta}_i^\top \boldsymbol{\mu}_i + \sum_{(i,j)\in\mathcal{E}} \boldsymbol{\theta}_{ij}^\top \boldsymbol{\mu}_{ij}.$$

$$\text{s.t.} \quad \boldsymbol{\mu}_{ij} = \text{vec}(\boldsymbol{\mu}_i \boldsymbol{\mu}_j^\top)$$

Combined LP and QP Relaxation

$$\min_{\boldsymbol{\mu}\in\mathcal{L}_\mathcal{G}} \boldsymbol{\theta}^\top \boldsymbol{\mu} + \rho g(\boldsymbol{\mu})$$

- $g(\boldsymbol{\mu})$: penalizes inconsistencies between $\boldsymbol{\mu}_{ij}$ and $\boldsymbol{\mu}_i \boldsymbol{\mu}_j^\top$.
- $\rho = 0$ amounts to the LP relaxation, $\rho \to \infty$ recovers the QP.
- Non-convex objective (for $\rho > 0$).
- Successively increase ρ to moderate the constraint enforcement.

Kullback-Leibler Divergence Penalty Terms – Two Variants

Uniform

$$\sum_{(i,j)\in\mathcal{E}} D_{KL}(\boldsymbol{\mu}_{ij}, \boldsymbol{\mu}_i \boldsymbol{\mu}_j^\top)$$

Tree-based for trees $\mathcal{G}_a = (\mathcal{V}_a, \mathcal{E}_a)$

$$\sum_{a\in\mathcal{A}} \eta_a \left(\sum_{(i,j)\in\mathcal{E}_a} D_{KL}(\boldsymbol{\mu}_{ij}, \boldsymbol{\mu}_i \boldsymbol{\mu}_j^\top) \right)$$

With $D_{KL}(\cdot)$ the Kullback-Leibler divergence or mutual information

$$D_{KL}(\boldsymbol{\mu}_{ij}, \boldsymbol{\mu}_i \boldsymbol{\mu}_j^\top) = \sum_{k,l} \mu_{ij;kl} \log \left(\frac{\mu_{ij;kl}}{\mu_{i;k}\mu_{j;l}} \right) = H(\boldsymbol{\mu}_i) + H(\boldsymbol{\mu}_j) - H(\boldsymbol{\mu}_{ij}).$$

With $H(\mathbf{p}) = -\sum_k p_k \log(p_k)$ the entropy of $\mathbf{p}$.

Concave-Convex Procedure (CCCP)

1. Decompose non-convex objective into convex u, v:

$$\min_{\boldsymbol{\mu}\in\mathcal{L}_\mathcal{G}} u_\rho(\boldsymbol{\mu}) - v_\rho(\boldsymbol{\mu})$$

2. Iteratively solve problem where v is linearized:

$$\boldsymbol{\mu}^{t+1} = \operatorname{argmin}_{\boldsymbol{\mu}\in\mathcal{L}_\mathcal{G}} u_\rho(\boldsymbol{\mu}) - \boldsymbol{\mu}^\top \nabla v_\rho(\boldsymbol{\mu}^t)$$

Difference of Convex Functions for Combined Relaxation

- Several solvers for LP relaxations add entropy terms to objective.
  $\to$ better convergence properties.
- *LPQP has an entropy augmentation built-in!*
- Linearized CCCP term simply modifies the unary potentials.

Uniform weighting

$$u_\rho(\boldsymbol{\mu}) = \boldsymbol{\theta}^\top \boldsymbol{\mu} - \rho \sum_{(i,j)\in\mathcal{E}} H(\boldsymbol{\mu}_{ij})$$

$$v_\rho(\boldsymbol{\mu}) = -\rho \sum_{i\in\mathcal{V}} d_i H(\boldsymbol{\mu}_i)$$

$$\frac{\partial v_\rho(\boldsymbol{\mu})}{\partial \mu_{i;k}} = \rho d_i (1 + \log \mu_{i;k})$$

Tree-based weighting

$$u_\rho(\boldsymbol{\mu}) = \boldsymbol{\theta}^\top \boldsymbol{\mu} - \rho \sum_{a\in\mathcal{A}} \eta_a \left(\sum_{(i,j)\in\mathcal{E}_a} H(\boldsymbol{\mu}_{ij}) - \sum_{i\in\mathcal{V}_a} (d_i^a - 1) H(\boldsymbol{\mu}_i) \right)$$

$$v_\rho(\boldsymbol{\mu}) = -\rho \sum_{a\in\mathcal{A}} \eta_a \sum_{i\in\mathcal{V}_a} H(\boldsymbol{\mu}_i)$$

$$\frac{\partial v_\rho(\boldsymbol{\mu})}{\partial \mu_{i;k}} = \rho \sum_{a\in\mathcal{A}(i)} \eta_a (1 + \log \mu_{i;k})$$

d_i : degree of node i in $\mathcal{G}$, d_i^a: degree of node i in graph $\mathcal{G}_a$, η_a: weight of graph $\mathcal{G}_a$.

LPQP Algorithm

Require: $\mathcal{G} = (\mathcal{V}, \mathcal{E}), \boldsymbol{\theta}$.

1: initialize $\boldsymbol{\mu} \in \mathcal{L}_\mathcal{G}$ uniform, $\rho = \rho_0$.
2: **repeat**
3: $t = 0$, $\boldsymbol{\mu}^0 = \boldsymbol{\mu}$.
4: **repeat**
5: $\boldsymbol{\mu}^{t+1} = \operatorname{argmin}_{\boldsymbol{\tau}\in\mathcal{L}_\mathcal{G}} u_\rho(\boldsymbol{\tau}) - \boldsymbol{\tau}^\top \nabla v_\rho(\boldsymbol{\mu}^t)$.
6: $t = t + 1$.
7: **until** $\|\boldsymbol{\mu}^t - \boldsymbol{\mu}^{t-1}\|_2 \leq \epsilon_{\text{dc}}$.
8: $\boldsymbol{\mu} = \boldsymbol{\mu}^t$.
9: increase ρ.
10: **until** $\|\boldsymbol{\mu} - \boldsymbol{\mu}^0\|_2 \leq \epsilon_\rho$.
11: **return** $\boldsymbol{\mu}$.

Inner Optimization Problem – Message Passing

For a fixed ρ and modified unary potentials $\tilde{\boldsymbol{\theta}}_i(\boldsymbol{\mu}_i^t)$ need to solve a convex optimization problem.

- Uniform weighting (LPQP-U):

$$\boldsymbol{\mu}^{t+1} = \operatorname{argmin}_{\boldsymbol{\mu}\in\mathcal{L}_\mathcal{G}} \tilde{\boldsymbol{\theta}}^\top \boldsymbol{\mu} - \rho \sum_{(i,j)\in\mathcal{E}} H(\boldsymbol{\mu}_{ij}).$$

  $\to$ Solved by norm-product belief-propagation.
- Tree-based weighting (LPQP-T):

$$\boldsymbol{\mu}^{t+1} = \operatorname{argmin}_{\boldsymbol{\mu}\in\mathcal{L}_\mathcal{G}} \tilde{\boldsymbol{\theta}}^\top \boldsymbol{\mu} - \rho \sum_{a\in\mathcal{A}} \eta_a \left(\sum_{(i,j)\in\mathcal{E}_a} H(\boldsymbol{\mu}_{ij}) - \sum_{i\in\mathcal{V}_a} (d_i^a - 1) H(\boldsymbol{\mu}_i) \right).$$

  $\to$ Solved by dual decomposition & the sum-product algorithm.

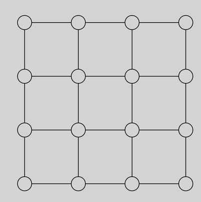
Synthetic Potts Model Experiment

- Potts model for a lattice graph.
- Mixed attractive and repulsive potentials.
- Lower signal-to-noise ratio for smaller σ.

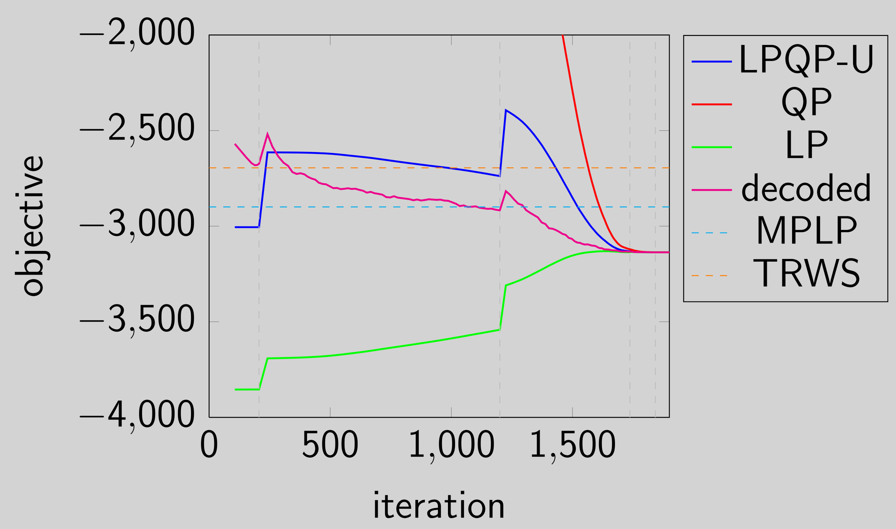
$$\theta_{i;k}(x_i) \sim \mathit{Uniform}(-\sigma, \sigma)$$

$$\alpha_{ij} \sim \mathit{Uniform}(-1, 1)$$

$$\theta_{ij}(x_i, x_j) = \begin{cases} 0 & \text{if } x_i \neq x_j \\ \alpha_{ij} & \text{otherwise.} \end{cases}$$



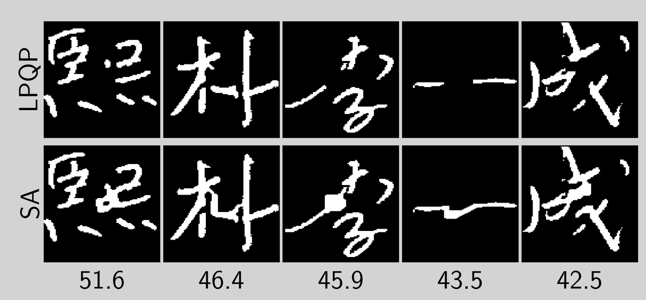
M (size)	60		90		120	
K (# states)	2	5	2	5	2	5
	$\sigma = 0.05$					
MPLP	0.71	0.99	0.51	0.96	0	0.95
LPQP-U	0.97	0.99	0.97	1	0.98	1
LPQP-T	1	0.97	1	0.98	1	0.98
TRWS	0	0	0	0	0.39	0
	$\sigma = 0.5$					
MPLP	1	1	1	1	1	0.99
LPQP-U	0.99	0.92	0.99	0.91	1	0.94
LPQP-T	0.99	0.95	0.99	0.94	0.99	0.96
TRWS	0	0	0	0	0	0



score for the evaluation: $s_i = \frac{\max_{1\leq j\leq J}(e_j) - e_i}{\max_{1\leq j\leq J}(e_j) - \min_{1\leq j\leq J}(e_j)}$

Real-world Experiments (Proteins and Inpainting)

DTF Problem Dataset



Better solution for 43/100 instances

Protein Prediction & Design

- Prediction. 28 instances where LP not tight. Recover true MAP for 2/3.
- Design. Average scores: LPQP-U: 0.93, MPLP: 1 and TRWS: 0.03.

Conclusions

- Joint formulation of LP and QP relaxation $\to$ LPQP.
- LPQP solved by a message-passing algorithm for modified unary potentials.
- Get a smooth objective for free. Key to fast convergence.
- Competitive results in terms of MAP state found.

References

- V. Kolmogorov (2006). "Convergent tree-reweighted message passing for energy minimization." In: *PAMI*
- Martin J Wainwright and Michael I Jordan (2008). *Graphical Models, Exponential Families, and Variational Inference*. Vol. 1
- A L Yuille and Anand Rangarajan (2003). "The concave-convex procedure". In: *Neural Computation*
- Akshat Kumar and Shlomo Zilberstein (2011). "Message-Passing Algorithms for Quadratic Programming Formulations of MAP Estimation". In: *UAI*
- Tamir Hazan and Amnon Shashua (2010). "Norm-Product Belief Propagation: Primal-Dual Message-Passing for Approximate Inference". In: *IEEE Transactions on Information Theory* 56.12, pp. 6294–6316
- David Sontag et al. (2008). "Tightening LP Relaxations for MAP using Message-Passing". In: *UAI*