A Combined LP and QP Relaxation for MAP

Patrick Pletscher, Sharon Wulff

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Energy Minimization and MAP inference

For a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,

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

is the discrete pairwise energy minimization problem

This problem:

- In general graphs and energies is NP-hard
- Arises in the context of finding the MAP prediction in MRF

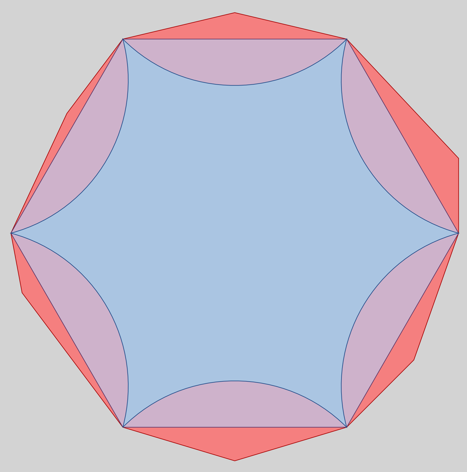
The potentials $\theta_i(x_i)$ and $\theta_{ij}(x_i, x_j)$ encode unary and pairwise dependencies

Relaxation Approaches

MAP as an integer quadratic program:

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

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



$\boldsymbol{\Theta}_{ij}$ represent the pairwise potentials in a matrix form.

Linear program (LP) relaxation:

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

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

Quadratic program (QP) relaxation:

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

$\Delta_K^{|\mathcal{V}|}$ is a product space of $|\mathcal{V}|$ simplexes over K variables

Combined LP and QP Relaxation

$$\min_{\boldsymbol{\mu} \in \mathcal{L}_\mathcal{G}} \underbrace{\sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^\mathsf{T} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^\mathsf{T} \boldsymbol{\mu}_{ij}}_{\text{standard LP objective}} + \beta \underbrace{\sum_{(i,j) \in \mathcal{E}} D_{KL}(\boldsymbol{\mu}_{ij}, \text{vec}(\boldsymbol{\mu}_i \boldsymbol{\mu}_j^\mathsf{T}))}_{\text{penalty term for inconsistencies}}$$

$$D_{KL}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \sum_k \mu_k \log\left(\frac{\mu_k}{\nu_k}\right) \quad \text{is the Kullback-Leibler divergence}$$

- While $\beta = 0$ amounts to the LP relaxation, $\beta \to \infty$ recovers the QP
- Option: successively increase β to moderate the constraint enforcement

Convex-Concave Procedure (CCCP)

1. Decompose non-convex objective into convex u, v:
$$\min_{\boldsymbol{\mu} \in \mathcal{L}_\mathcal{G}} u_\beta(\boldsymbol{\mu}) - v_\beta(\boldsymbol{\mu})$$
2. Iteratively solve problem where v is linearized:
$$\boldsymbol{\mu}^{t+1} = \operatorname*{argmin}_{\boldsymbol{\mu} \in \mathcal{L}_\mathcal{G}} u_\beta(\boldsymbol{\mu}) - \boldsymbol{\mu}^\mathsf{T} \nabla v_\beta(\boldsymbol{\mu}^t)$$

CCCP for the Combined LP-QP Relaxation

$$u_\beta(\boldsymbol{\mu}) = \sum_{i \in \mathcal{V}} \boldsymbol{\theta}_i^\mathsf{T} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^\mathsf{T} \boldsymbol{\mu}_{ij} - \beta \sum_{(i,j) \in \mathcal{E}} H(\boldsymbol{\mu}_{ij})$$

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

d_i: degree of node i
$H(\boldsymbol{\mu})$: entropy of $\boldsymbol{\mu}$

The gradient of $v_\beta(\boldsymbol{\mu})$:

$$\frac{\partial v_\beta(\boldsymbol{\mu})}{\partial \mu_{i;k}} = \beta d_i (1 + \log(\mu_{i;k})) \qquad \text{and} \qquad \frac{\partial v_\beta(\boldsymbol{\mu})}{\partial \mu_{ij;kl}} = 0$$

LPQP Algorithm

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

1: initialize $\boldsymbol{\mu} \in \mathcal{L}_\mathcal{G}$ uniform, $\beta = \beta_0$.
2: **repeat**
3: $t = 0, \boldsymbol{\mu}^0 = \boldsymbol{\mu}$.
4: **repeat**
5: $\boldsymbol{\mu}^{t+1} = \text{argmin}_{\boldsymbol{\tau} \in \mathcal{L}_\mathcal{G}} u_\beta(\boldsymbol{\tau}) - \boldsymbol{\tau}^\mathsf{T} \nabla v_\beta(\boldsymbol{\mu}^t)$.
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 β.
10: **until** $\|\boldsymbol{\mu} - \boldsymbol{\mu}^0\|_2 \leq \epsilon_\beta$.
11: **return** $\boldsymbol{\mu}$.

Inner Optimization Problem – Message Passing

For a fixed β and $\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i - \beta d_i \log(\boldsymbol{\mu}_i^t)$ we solve,

$$\boldsymbol{\mu}^{t+1} = \operatorname*{argmin}_{\boldsymbol{\mu} \in \mathcal{L}_\mathcal{G}} \sum_{i \in \mathcal{V}} \tilde{\boldsymbol{\theta}}_i^\mathsf{T} \boldsymbol{\mu}_i + \sum_{(i,j) \in \mathcal{E}} \boldsymbol{\theta}_{ij}^\mathsf{T} \boldsymbol{\mu}_{ij} - \beta \sum_{(i,j) \in \mathcal{E}} H(\boldsymbol{\mu}_{ij})$$

Using a variant of the norm-product belief propagation algorithm

$$m_{j \to i}(x_i) \propto \left(\sum_{x_j} \theta_{ij}^{1/\beta}(x_i, x_j) \frac{\theta_j^{1/(d_j\beta)}(x_j) \prod_{k \in \mathcal{N}(j)} m_{k \to j}^{1/(d_j\beta)}(x_j)}{m_{i \to j}^{1/\beta}(x_j)} \right)^\beta$$

$$\mu_i(x_i) \propto \left(-\theta_i(x_i) \prod_{j \in \mathcal{N}(i)} m_{j \to i}(x_i) \right)^{1/(d_i\beta)}$$

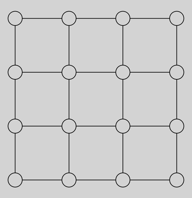
Synthetic Lattice Experiment

- Potts model for a lattice graph.
- Mixed attractive and repulsive potentials.
- For smaller σ: Unaries have less influence.

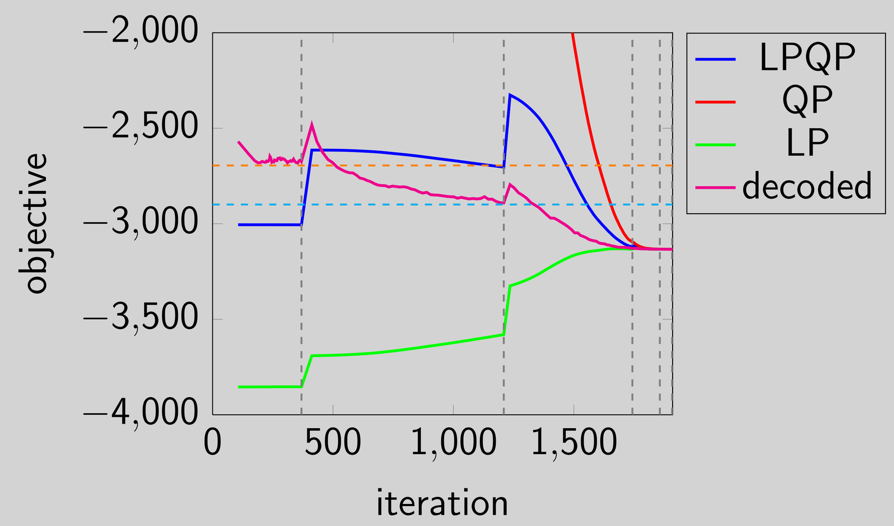
$$\theta_{i;k}(x_i) \sim \mathit{Uniform}(-\sigma, \sigma)$$

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

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

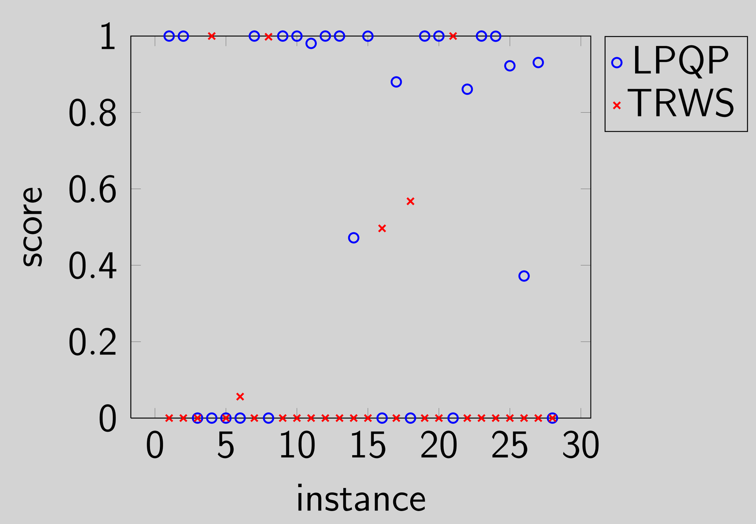


M (size)	60		90		120	
K (# states)	2	5	2	5	2	5
			$\sigma = 0.05$			
MPLP	0.73	0.99	0.53	0.97	0	0.95
LPQP	1	0.99	1	1	1	1
TRWS	0	0	0	0	0.40	0
			$\sigma = 0.5$			
MPLP	1	1	1	1	1	1
LPQP	0.99	0.94	0.99	0.93	1	0.96
TRWS	0	0	0	0	0	0



Protein Side-Chain Prediction Experiment

- 28 instances where LP not tight.
- True MAP state known through integer programming.
- LPQP only on six instances worse than TRWS.
- LPQP finds global minimum for 12 instances.



Conclusions

- Joint formulation of LP and QP relaxation.
- CCCP algorithm in its core uses norm-product belief-propagation for modified unary potentials.
- Competitive results in terms of MAP state found.

References

- Pradeep Ravikumar and John Lafferty (2006). "Quadratic Programming Relaxations for Metric Labeling and Markov Random Field MAP Estimation". In: *ICML*
- 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*