Graph Preserving Label Decomposition in Discrete MRFs with Selfish Potentials

Dmitry Vetrov* **Anton Osokin***
Department of Computational Mathematics and Cybernetics
Moscow State University
vetrovd@yandex.ru anton.osokin@gmail.com

Abstract

In this paper we address the problem of finding the most probable state of discrete Markov random field (MRF) with pairwise terms of a special type which we refer to as "selfish". Selfish potentials allow us to use a novel type of MRF decomposition to a number of solvable subproblems with the same set of nodes and edges but binary variables. We call the proposed framework Graph Preserving Label Decomposition (GPLD) and prove that the GPLD lower bound is equal to the solution of the standard LP-relaxation of the initial problem. The special structure of GPLD makes it possible to take into account the desired global properties of the solution, e.g. constraints on various linear combinations of class indicator variables. We provide the comparison of our method with state-of-the-art algorithms both in terms of accuracy and speed.

1 Introduction

The problem of efficient Bayesian inference arises in many applied domains, e.g.in machine learning, computer vision, decision-making, etc. One of the most intriguing problems is the development of approximate inference algorithms for problems that are NP-hard in general. An important particular case is the MAP-inference problem in cyclic discrete Markov random fields (MRF) with energies that can be represented via sum of unary and pairwise terms.

Let $G = (\mathcal{V}, \mathcal{E})$ be an undirected graph with $\mathcal{V}$ and $\mathcal{E}$ being the sets of nodes and edges, respectively. With each node we associate a class label $t_j \in \{1, \dots, P\}$. The MAP-inference problem can be formulated as an energy minimization problem

$$\sum_{(j \in \mathcal{V})} \theta_j(t_j) + \sum_{(i,j) \in \mathcal{E}} \theta_{ij}(t_i, t_j) \to \min_{t_1, \dots, t_{|\mathcal{V}|} \in \{1, \dots, P\}}. \tag{1}$$

where unary potentials $\theta_j(t_j)$ and pairwise potentials $\theta_{ij}(t_i, t_j)$ are some known functions of discrete argument.

Although NP-hard in general problem (1) can be solved exactly in polynomial time in several special cases. One example is dynamic programming approach [11] for inference in tree-structured graphs. Another example is MRFs with outer-planar graphs [13], or more generally MRFs with low treewidth [8]. Min-cut/max-flow algorithms can efficiently solve the MAP-inference problem on arbitrary graphs when all the variables are binary and the pairwise potentials meet submodularity constraint [4, 6].

$$\theta_{ij}(0,0) + \theta_{ij}(1,1) \leq \theta_{ij}(0,1) + \theta_{ij}(1,0). \tag{2}$$

*The authors assert equal contribution and thus joint first authorship.

When the submodularity property doesn't hold one can use quadratic pseudo-boolean optimization algorithm (QPBO) to get a lower bound on the minimum of energy (1). This lower bound is known to equal the solution of the LP-relaxation of the binary problem [12].

Advanced approximate methods based on MRF decomposition have recently appeared [14]. The most popular method, tree-reweighted message passing (TRW) [16, 5, 7], splits the MRF with cycles into a number of acyclic subgraphs (trees) and, for each tree, inference is made independently with the subsequent harmonization of the optimal solutions. Although being able to solve discrete subproblems exactly tree decomposition methods are known to converge to the solution of the LP-relaxation of the initial problem at their best[1]. The other decomposition methods [15, 1] exploit similar ideas. Unlike more efficient approximate energy minimization algorithms, e.g. [3], decomposition framework makes it possible to take into account some global properties of the solution, in particular to establish constraints on the areas of classes [17, 9].

In this paper we address the MAP-inference problem with pairwise potentials of a special kind which we call "selfish" since they care (either attract or repulse) the neighbors of the same class returning zero otherwise. Selfish potentials allow us to perform the decomposition to a number of binary subproblems that correspond to different classes and are based on the set of nodes $\mathcal{V}$ and the set of edges $\mathcal{E}$. These subproblems are NP-hard in general but we can use QPBO algorithm to get the solution of their LP-relaxation. We prove that the harmonization of subproblems' solutions via dual decomposition provides the LP-relaxation of the initial problem. We show that our algorithm generally converges faster than TRW algorithm which is based on the dual decomposition (DD-TRW) and provides better lower bound that TRW-S in the presence of repulsive pairwise potentials. We refer to the new type of decomposition as graph preserving label decomposition (GPLD). GPLD also allows us to take into account the preferences on any type of global linear statistics of the class indicator variables in a straightforward manner.

The rest of the paper is organized as follows. In the next section we present GPLD framework and prove the equivalence of GPLD and LP-relaxation lower bounds. The way of taking into account some global conditions on the desired solution is discussed in section 3. We present some experimental results in section 4.

2 Graph Preserving Label Decomposition

2.1 Decomposition of standard MRFs with selfish pairwise potentials

Consider the indicator parametrization of (1) obtained by establishing auxiliary binary variables $Y = \{y_{jp}\} \in \{0,1\}^{|\mathcal{V}|\times P}$:

$$y_{jp} = \begin{cases} 1, & t_j = p, \\ 0, & \text{otherwise.} \end{cases}$$

Denote $\theta_j(p) = \theta_{jp}$ and $\theta_{ij}(p,q) = \theta_{ij,pq} = \theta_{ji,qp}$. Then problem (1) takes the form of

$$E(Y) = \sum_{j\in\mathcal{V}}\sum_{p=1}^{P}\theta_{jp}y_{jp} + \sum_{(i,j)\in\mathcal{E}}\sum_{p,q=1}^{P}\theta_{ij,pq}y_{ip}y_{jq} \to \min_{Y}, \tag{3}$$

$$\text{s.t. } y_{jp} \in \{0,1\},\ \sum_{p=1}^{P} y_{jp} = 1. \tag{4}$$

We denote set $\left\{Y \mid y_{jp} \in \{0,1\}\right\}$ by $\mathcal{L}$ and set $\left\{Y \mid \sum_{p=1}^{P} y_{jp} = 1,\ \forall j \in \mathcal{V}\right\}$ by $\mathcal{G}$.

In what follows we consider selfish pairwise potentials, i.e. such ones that

$$\theta_{ij,pq} = C_{ij,p}\delta_{pq}, \tag{5}$$

[1]The popular TRW-S algorithm does not have this property and guarantees the convergence only to a coordinate-wise maximum of the lower bound

where $\delta_{pq} = 1$ iff $p = q$. Then we may rewrite energy (3) as follows

$$\min_{Y \in \mathcal{L} \cap \mathcal{G}} E(Y) = \min_{Y \in \mathcal{L} \cap \mathcal{G}} \sum_{p=1}^{P} \left(\sum_{j \in \mathcal{V}} \theta_{jp} y_{jp} + \sum_{(i,j) \in \mathcal{E}} C_{ij,p} y_{ip} y_{jp} \right) \geq$$

$$\max_{\Lambda} \min_{Y \in \mathcal{L}} \sum_{p=1}^{P} \left(\sum_{j \in \mathcal{V}} \theta_{jp} y_{jp} + \sum_{(i,j) \in \mathcal{E}} C_{ij,p} y_{ip} y_{jp} + \sum_{j \in \mathcal{V}} \lambda_j \left(\sum_{p=1}^{P} y_{jp} - 1 \right) \right) =$$

$$\max_{\Lambda} \left\{ \sum_{p=1}^{P} \min_{Y \in \mathcal{L}} \left(\sum_{j \in \mathcal{V}} (\theta_{jp} + \lambda_j) y_{jp} + \sum_{(i,j) \in \mathcal{E}} C_{ij,p} y_{ip} y_{jp} \right) - \sum_{j \in \mathcal{V}} \lambda_j \right\}. \quad (6)$$

Hence we obtained a decomposition of the initial problem to P subproblems each corresponding to a single class.

2.2 Reduction to solvable subproblems

Note that if all parameters $C_{ij,p}$ were non-positive the subproblems of (6) could be solved easily using min-cut algorithms since their energies would be submodular. This forms the basis of recently proposed submodular decomposition method (SMD) [10]. However in general case these subproblems are still NP-hard. Consider a single subproblem which corresponds to label p. We may rewrite it as a linear function by adding new variables Z

$$\min_{Y \in \mathcal{L}} \left(\sum_{j \in \mathcal{V}} (\theta_{jp} + \lambda_j) y_{jp} + \sum_{(i,j) \in \mathcal{E}} C_{ij,p} y_{ip} y_{jp} \right) = \min_{Y,Z} \left(\sum_{j \in \mathcal{V}} (\theta_{jp} + \lambda_j) y_{jp} + \sum_{(i,j) \in \mathcal{E}} C_{ij,p} z_{ij,pp} \right) \quad (7)$$

$$\text{s.t. } y_{jp}, z_{ij,pp} \in \{0,1\}, \;\; z_{ij,pp} \leq y_{ip}, y_{jp}, \;\; z_{ij,pp} \geq y_{ip} + y_{jp} - 1. \quad (8)$$

Removal of integrality constraints with non-negativity conditions on Y, Z yields linear programming problem.

$$\sum_{j \in \mathcal{V}} (\theta_{jp} + \lambda_j) y_{jp} + \sum_{(i,j) \in \mathcal{E}} C_{ij,p} z_{ij,pp} \to \min_{Y,Z} \quad (9)$$

$$\text{s.t. } y_{jp}, z_{ij,pp} \in [0,1], \;\; z_{ij,pp} \geq 0, \;\; z_{ij,pp} \leq y_{ip}, y_{jp}, \;\; z_{ij,pp} \geq y_{ip} + y_{jp} - 1. \quad (10)$$

Lemma 1 *Problem* (9) *with the constraints* (10) *is equivalent to problem* (9) *with the following constraints*

$$y_{ip} + y_{i\overline{p}} = 1, \; y_{ip}, y_{i\overline{p}}, z_{ij,pp}, z_{ij,\overline{p}p}, z_{ij,p\overline{p}}, z_{ij,\overline{pp}} \geq 0, \quad (11)$$

$$z_{ij,p\overline{p}} + z_{ij,pp} = y_{ip}, \; z_{ij,p\overline{p}} + z_{ij,\overline{pp}} = y_{i\overline{p}}, \; z_{ij,\overline{p}p} + z_{ij,pp} = y_{jp}, \; z_{ij,\overline{pp}} + z_{ij,p\overline{p}} = y_{j\overline{p}} \quad (12)$$

The proof requires consideration of the sign of $C_{ij,p}$. In the case of negative $C_{ij,p}$ the constraints (11), (12) imply $z_{ij,pp} \leq \min(y_{ip}, y_{jp})$. In the case of positive $C_{ij,p}$ the constraints (11), (12) imply $z_{ij,pp} \geq y_{ip} + y_{jp} - 1$. On the other hand after finding the solution of (9), (10) we may always set $z_{ij,\overline{p}p} = y_{jp} - z_{ijpp}$, $z_{ij,p\overline{p}} = y_{ip} - z_{ij,pp}$, $z_{ij,\overline{pp}} = z_{ij,pp} + 1 - y_{ip} - y_{jp}$. It is straightforward to show that all constraints (11), (12) are satisfied.

Note that minimum of (9) w.r.t. (11), (12) can be found efficiently using quadratic pseudo-Boolean optimization (QPBO). Formally QPBO computes only the value of minimum and assigns either zero, one, or "don't know" to y_{ip}. However it can be shown [2] that "don't know" answers correspond to $y_{ip} = 0.5$ (so-called half-intergrality property) in the solution of (9) and the missing values of $z_{ij,pp}$ can be easily revealed using lemma 1 and (10):

$$z_{ij,pp} = \begin{cases} \min(y_{ip}, y_{jp}), & C_{ij,p} \leq 0, \\ \max(0, y_{ip} + y_{jp} - 1), & \text{otherwise.} \end{cases} \quad (13)$$

Finally we may establish a lower bound on lower bound (6)

$$\max_{\Lambda}\left\{\sum_{p=1}^{P}\min_{Y\in\mathcal{L}}\left(\sum_{j\in\mathcal{V}}(\theta_{jp}+\lambda_j)y_{jp}+\sum_{(i,j)\in\mathcal{E}}C_{ij,p}y_{ip}y_{jp}\right)-\sum_{j\in\mathcal{V}}\lambda_j\right\}\geq$$

$$\max_{\Lambda}\left\{\sum_{p=1}^{P}\min_{Y,Z\in(10)}\left(\sum_{j\in\mathcal{V}}(\theta_{jp}+\lambda_j)y_{jp}+\sum_{(i,j)\in\mathcal{E}}C_{ij,p}z_{ij,pp}\right)-\sum_{j\in\mathcal{V}}\lambda_j\right\}=$$

$$\max_{\Lambda}\left\{\min_{Y,Z\in(10)}E(Y,Z,\Lambda)-\sum_{j\in\mathcal{V}}\lambda_j\right\}=\max_{\Lambda}\Phi(\Lambda) \quad (14)$$

where each subproblem is solvable. Note that with respect to Λ lower bound (14) is piecewise linear concave function and hence can be maximized e.g. by subgradient ascent.

2.3 Convergence to LP relaxation bound

GPLD allows to get a lower bound on the solution of discrete problem (3), (4). In SMD it was possible to optimize directly (6) and it was proven that its maximum equals to the solution of the LP-relaxation of the initial problem

$$\sum_{j\in\mathcal{V}}\sum_{p=1}^{K}\theta_{jp}y_{jp}+\sum_{(i,j)\in\mathcal{E}}\sum_{p,q=1}^{K}\theta_{ij,pq}z_{ij,pq}\to\min_{Y,Z} \quad (15)$$

$$\text{s.t. } y_{jp},z_{ij,pq}\geq 0,\ \sum_{p=1}^{K}y_{jp}=1,\ \sum_{q=1}^{K}z_{ij,pq}=y_{ip},\ \sum_{p=1}^{K}z_{ij,pq}=y_{jq}. \quad (16)$$

In GPLD one can optimize weaker lower bound (14). Nevertheless the following statement holds.

Theorem 1 *The maximum of GPLD lower bound* (14) *equals to the minimum of LP-relaxation of energy* (15)*,* (16)*.*

Due to the lack of space we provide just a short sketch of the proof. Consider $\Lambda^0=\arg\max\Phi(\Lambda)$. Among the argmins of $E(Y,Z,\Lambda^0)$ there exists (Y^0,Z^0) such that $Y^0\in\mathcal{G}$. It suffices to show that we may always find $Z^1\geq 0$ such that all $z^1_{ij,pp}$ satisfy (13), $\sum_{q=1}^{K}z^1_{ij,pq}=y^0_{ip}$, $\sum_{p=1}^{K}z^1_{ij,pq}=y^0_{jq}$, and $E(Y^0,Z^0,\Lambda^0)=E(Y^0,Z^1,\Lambda^0)$. In the proof we provide a constructive algorithm that builds such Z^1.

3 Global constraints

The graph preserving label decomposition has several advantages over state-of-the art (wide)tree-based decompositions. Since we deal with label indicator variables y_{jp} we may establish any kinds of constraints on linear functions of these variables

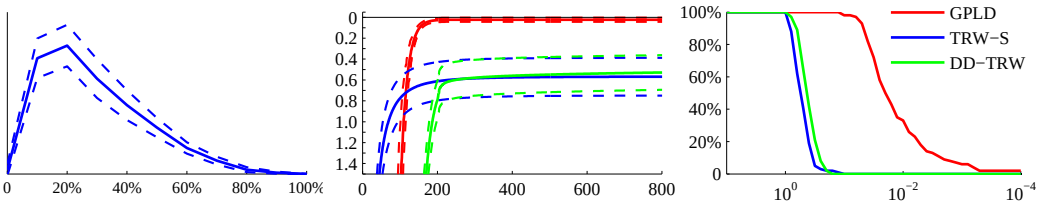
$$\sum_{j\in\mathcal{V}}\sum_{p=1}^{P}w^m_{jp}y_{jp}=c^m,\ \ m=1,\dots,M \quad (17)$$

$$\sum_{j\in\mathcal{V}}\sum_{p=1}^{P}v^k_{jp}y_{jp}\leq d^k,\ \ k=1,\dots,K. \quad (18)$$

Then p^{th} subproblem takes the form of

$$\min_{Y,Z\in(10)}\left(\sum_{j\in\mathcal{V}}(\theta_{jp}+\lambda_j+\sum_{m=1}^{M}\mu_m w^m_{jp}+\sum_{k=1}^{K}\kappa_k v^k_{jp})y_{jp}+\sum_{(i,j)\in\mathcal{E}}C_{ij,p}z_{ij,pp}\right), \quad (19)$$

where Λ, M, and $K\geq 0$ are Lagrange multipliers with respect to which one performs further projected subgradient ascend. Note that the subproblem is still solvable by QPBO.



(a) (b) (c)

Figure 1: (a) shows the estimate of the duality gap against the rate of repulsive pairwise potentials. Solid line shows the mean value, dashed lines show one standard deviation from the mean value. (b) demonstrates the average lower bound on energy 1 obtained by GPLD (red), TRW-S (blue), DD-TRW (green) vs. number of iterations. The vertical axis indicates the gap between the current lower bound and the solution of the LP-relaxation. The value of the gap between the integer solution obtained by TRW-S and the exact solution of the LP-relaxation obtained using an interior point solver is set to 100. The horizontal axis shows the current iteration number. (c) shows the percentage of generated problems against the accuracy (the difference between the lower bound and the LP-relaxation) achieved within 5000 iterations. The horizontal axis is in the logarithmic scale and its units correspond to the units of the vertical axis of plot (b).

Theorem 2 *Maximin*

$$\max_{\Lambda,M,K\geq 0}\left(\min_{Y,Z\in(10)} E(Y,Z,\Lambda,M,K) - \sum_{j\in\mathcal{V}}\left(\lambda_j + \sum_{m=1}^{M} c_m\mu_m + \sum_{k=1}^{K} d_k\kappa_k\right)\right)$$

equals to the solution of LP problem (15), (16) *with additional constraints* (17), (18).

The proof is similar to the proof of theorem 1. The only difference is that Y^0 also satisfies (17) and (18).

4 Experiments and discussion

Evaluation of the primal-dual gap w.r.t. the rate of non-submodular terms. The "complexity" of problem (1) greatly depends on the relations between energy parameters. It has been shown (e.g. [5]) that in the presence of both attractive and repulsive pairwise potentials the duality gap (and consequently the "complexity" of the problem) grows with the increase of relative strength of pairwise potentials. Here we explore another important factor – the fraction of repulsive Potts potentials, i.e. the ones where $C_{ij,p} > 0$. Figure 1a shows the estimate of the duality gap against the rate of repulsive potentials to all pairwise potentials. To ensure the reproductivity of this result we estimate the duality gap using TRWS-S method[2]. For each rate of repulsive potentials we generate 20 toy problems: 50×50 grid, 10 labels, unary potentials are generated from $\mathcal{N}(0,1)$, pairwise potentials are generated as absolute values of $\mathcal{N}(0,2)$ with subsequent sign switch of the given fraction of randomly chosen potentials. Figure 1a shows that the largest duality gap appears when 20-30% of pairwise potentials are repulsive.

Comparison with TRW-S and DD-TRW. In this section we compare the performance of TRW-S, DD-TRW, and GPLD on a set of 100 syntectic problems generated in the same setup described in the previous section. The rate of repulsive potentials is kept at the level of 30%. For TRW-S method we use the author's original code. In DD-TRW we decompose the grid into vertical and horizontal chains. As for subgradient optimization both in DD-TRW and GPLD we've selected an adaptive scheme that was recommended in [7, eq. 41] combined with further switch to non-adaptive diminishing step size rule: $\alpha_t = \frac{a}{\sqrt{t}}$ that gives theoretical guarantees of the convergence to the optimum. Here positive constant a is a parameter of the algorithm and t is the current iteration number.

Plot 1b shows the averaged performance of GPLD, TRW-S, and DD-TRW against the number of iterations. Plot 1c shows the fraction of problems where the given method achieved the given level of accuracy within 5000 iterations. We observe that TRW-S shows the best performance in the beginning but gets stuck in a coordinate-wise maximum, i.e. is generally unable to converge to

[2]We used the original authors code published on http://pub.ist.ac.at/ vnk/papers/TRW-S.html

(a) (b) (c) (d)

Figure 2: Results for the image segmentation problem. (a) – initial image with seeds; (b) – the result of TRW-S; (c) – result of GPLD; (d) – the result of GPLD with global constraints on label areas.

the LP lower bound. GPLD shows the ability to achieve higher accuracy in reasonable time. Poor performance of DD-TRW can be explained by the fact that the dimensionality of the dual space is $P = 10$ times larger than the dimensionality of GPLD's dual space. Another possible explanation is that our way of decomposing the grid into trees may not optimal and implies slow convergence.

Image segmentation with non-submodular pairwise terms and global constraints. To show the effect of global linear constraints on indicator variables we construct an energy to segment artificial image 2a with user provided seeds. Identical colors of all 3 objects and background make color-based unary potentials useless and therefore unaries contain only seed information and small bias to background class. Pairwise potentials are set to attractive Potts potentials $C_{ij,p} = -C < 0$ at the edges with low contrast and to repulsive Potts potentials $C_{ij,p} = C$ at the edges with high contrast. The contrast threshold is set to 25 (black color has intensity 0 and white color has intensity 255). In this setting the energy is "complex", because the unary potentials are weak and the pairwise potentials are both repulsive and attractive. Figures 2b, c show that both TRW-S and GPLD cope with the problem poorly i.e. provide very fragmented segemtnation. Figure 2d shows the result of GPLD with global constraints on objects' areas which were set to their right values.

Conclusion. Our experiments show that GPLD outperforms DD-TRW in time and TRW-S in accuracy in the "complex" cases when there are both repulsive and attractive pairwise terms. A possible direction for the future work would be to apply probing algorithms [12] which yield tighter relexations for binary subproblems (7).

References

[1] D. Batra, A. C. Gallagher, D. Parikh, and T. Chen. Beyond trees: MRF inference via outer-planar decomposition. In *CVPR*, 2010.

[2] E. Boros and P. L. Hammer. Pseudo-boolean optimization. *Discrete Applied Mathematics*, 123(1-3):155–225, 2002.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast Approximate Energy Minimization via Graph Cuts. *PAMI*, 23(11):1222–1239, 2001.

[4] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2):271–279, 1989.

[5] V. Kolmogorov. Convergent Tree-Reweighted Message Passing for Energy Minimization. *PAMI*, 28(10):1568–1583, 2006.

[6] V. Kolmogorov and R. Zabih. What Energy Functions Can Be Optimized via Graph Cuts. *PAMI*, 26(2):147–159, 2004.

[7] N. Komodakis, N. Paragios, and G. Tziritas. MRF Energy Minimization and Beyond via Dual Decomposition. *PAMI*, 33(3):531–552, 2010.

[8] S. L. Lauritzen. *Graphical models*. Clarendon Press, 1996.

[9] Y. Lim, K. Jung, and P. Kohli. Energy Minimization Under Constraints on Label Counts. In *ECCV*, 2010.

[10] A. Osokin, D. Vetrov, and V. Kolmogorov. Submodular decomposition framework for inference in associative markov networks with global constraints. In *CVPR*, 2011.

[11] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Francisco, 1988.

[12] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, 2007.

[13] N. N. Schraudolph and D. Kamenetsky. Efficient Exact Inference in Planar Ising Models. In *NIPS*, 2008.

[14] D. Sontag, A. Globerson, and T. Jaakkola. Introduction to dual decomposition for inference. In S. Sra, S. Nowozin, and S. J. Wright, editors, *Optimization for Machine Learning*. MIT Press, 2011.

[15] D. Sontag, T. Meltzer, A. Globerson, Y. Weiss, and T. Jaakkola. Tightening LP relaxations for MAP using message-passing. In *UAI*, 2008.

[16] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. MAP estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 51(11):3697–3717, 2005.

[17] O. J. Woodford, C. Rother, and V. Kolmogorov. A Global Perspective on MAP Inference for Low-Level Vision. In *ICCV*, 2009.