Planar Decompositions and Cycle Constraints

Julian Yarkony, Ragib Morshed, Alexander T. Ihler, Charless C. Fowlkes

Dept. of Computer Science, University of California, Irvine, CA 92697

{jyarkony,rmorshed,ihler,fowlkes}@ics.uci.edu

1. Introduction

Dual-decomposition methods for optimization have emerged as an extremely powerful tool for solving combinatorial problems in graphical models. These techniques can be thought of as decomposing a complex model into a collection of easier-to-solve components, providing a variational bound which can then be optimized over its parameters. A wide variety of algorithms have been proposed, often distinguished by the class of models from which subproblems are constructed, including trees [14, 5], planar graphs [3], outer-planar graphs [2], k-fans [4], or some more heterogeneous mix of combinatorial subproblems [13].

While the class of tree-reweighted methods are now fairly well understood, many of the same concepts and guidance available for trees are not available for more general classes of decompositions. In this work, we analyze reweighting methods that seek to decompose planar MRFs into subproblems consisting of tractable binary planar graphs. We show that the ultimate building blocks of such a decomposition are simple cycles of the original graph and that to achieve the tightest possible bounds, one must choose a set of subproblems that cover all such cycles. Cycles in planar-reweighted decomposition thus play a role analogous to trees in tree-reweighted decompositions.

There are various techniques for enforcing consistency over cycles in an MRF. For example, one can triangulate the graph and introduce constraints over all triplets in the resulting triangulation. However, this involves $O(n^3)$ constraints which is impractical in large-scale inference problems. A more efficient route is to only add a small number of constraints as needed, e.g., using a cutting-plane approach [11].

We describe a new variational bound that enforces the constraints over *all cycles* in a planar binary MRF with only a constant factor overhead. This representation is very simple and efficient to optimize, which we demonstrate in experimental comparisons to existing state-of-the-art, cycle-enforcing methods, where we achieve substantial performance gains. To handle more general d-state planar MRFs, we consider binary partitions of the state space of each node which yield tractable binary subproblems. There are a large number of such subproblems. However, we are often able to achieve tight bounds after adding only a handful of sub-

problems in an incremental fashion, analogous to cutting plane techniques used in LP relaxations, and outperform competing algorithms on classes of problems which have long range cyclic dependencies.

2. Exact Inference for Binary Planar MRFs

Consider the energy function $E(X)$ associated with a general binary MRF defined over a collection of variables $(X_1, X_2, \ldots) \in \{0, 1\}^N$ with specified unary and pairwise potentials. It is straightforward to show that any such MRF can be reparametrized up to a constant using pairwise disagreement costs θ_{ij} along with unary parameters θ_i. The energy function can thus be written as

$$E(X, \theta) = \sum_{i>j} \theta_{ij}[X_i \neq X_j] + \sum_i \theta_i[X_i \neq 0] \quad (1)$$

where $[\cdot]$ is the indicator function and we have dropped any constant terms.

We can express such an energy function without including any unary terms by introducing an auxiliary variable X_0 and replacing the unary terms with pairwise connections to X_0 so that

$$E_1(X, \theta) = \sum_{i>j} \theta_{ij}[X_i \neq X_j] + \sum_i \theta_i[X_i \neq X_0] \quad (2)$$

If we fix $X_0 = 0$, then E_1 is clearly equivalent to our original energy function E. Since the potentials in E_1 are symmetric, for any state $X = (X_0, X_1, \ldots)$, there is a state $\bar{X}$ with identical energy, given by flipping the states of every X_i including X_0. Thus any X that minimizes E_1 can be easily mapped to a minimizer of E.

While minimizing $E(X, \theta)$ is computationally intractable in general [1], a clever construction due to Kasteleyn and Fisher allows one to find minimizing states when the graph corresponding to E_1 is planar. This is based on the complementary relation between states of the nodes X and perfect matchings in the so-called expanded dual of the graph $\mathcal{G}_1$. [1] See the report of [9] for an in-depth discussion and implementation details.

---

[1]Matchings in planar graphs can be found somewhat more efficiently than for general graphs which yields the best known worst-case running time of $O(N^{3/2} \log N)$ for max-cut in planar graphs [10].

While this reduction to perfect matching provides a unique tool for energy minimization and probabilistic inference, the requirement that $\mathcal{G}_1$ be planar is a serious restriction since even if the original graph $\mathcal{G}$ corresponding to E is planar, e.g., in the case of the grid graphs commonly used in computer vision applications, $\mathcal{G}_1$ is typically not, since the addition of edges from every node to the auxiliary node X_0 renders the graph non-planar.

3. Inference with Dual Decomposition

Dual decomposition is a general approach for leveraging such islands of tractability in order to perform inference in more general MRFs. The application of dual decomposition to inference in graphical models was popularized by the work of [14] on Tree-Reweighted Belief Propagation (TRW). TRW finds an optimal decomposition of an MRF into a collection of tree-structured problems where exact inference is tractable. More formally, let t index a collection of subproblems defined over the same set of variables X and whose parameters sum up to the original parameter values, so that $\theta = \sum_t \theta^t$. The energy function is linear in θ so we have

$$E_{MAP} = \min_X E(X, \Theta) = \min_X \sum_t E(X, \Theta^t) \quad (3)$$

$$\geq \max_{\sum_t \theta^t = \theta} \sum_t \min_{X^t} E(X^t, \Theta^t) \quad (4)$$

The inequality arises because each subproblem t is solved independently and thus may yield different solutions. On the other hand, if the solutions to the sub-problems all happen to agree then the bound is tight. The problem of maximizing the lower-bound over possible decompositions $\{\theta^t\}$ is convex and when inference for each sub-problem is tractable (for example, θ^t is tree-structured) the bound can be optimized efficiently using message passing (fixed-point iterations) based on computing min-marginals in each subproblem [14] or by projected subgradient methods [7].

One can tighten the bound in Equation 4 by adding additional subproblems to the primal (or equivalently constraints to the dual) which enforce consistency over larger sets of variables. This has been explored, e.g. by [11] who suggest adding cycle inequalities to the dual which enforce consistency of pseudo-marginals around a cycle. Since there are a large number of potential cycles present in the graph, Sontag suggests either using a cutting plane algorithm to successively add violated cycle constraints [11] or to only add small cycles such as triplets or quadruplets [12] that can be enumerated with relative ease and optimized using local message passing rather than general LP solvers.

For binary problems, it is natural to consider replacing Wainwright's tree subproblems with tractable outer-planar subgraphs. This has been explored by [3] and [2] who proposed decomposing a graph into a set of planar graphs for
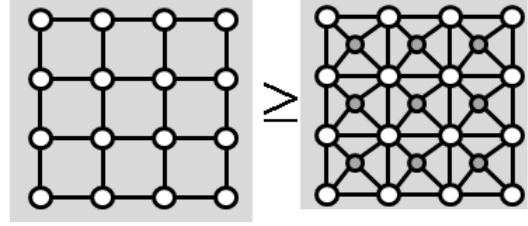

Figure 1. *We lower-bound the energy of a planar binary MRF by a tractable planar graph in which auxiliary nodes associated with each face capture unary potentials in the original problem.*

the purposes of estimating the partition function and minimum energy state respectively. For energy minimization, it is well-known that any set of subproblems that cover every edge is sufficient to achieve the TRW bound; but what is the best set of planar graphs to use? Is it necessary to use all outer-planar or even all planar subgraphs? It turns out that consistency on the set of all outer-planar or planar subgraphs is equivalent to consistency on all cycles of $\mathcal{G}$. This observation leads to algorithms such as reweighted perfect matching [8], which explicitly constructs a set of subproblems that form a complete cycle basis, or incremental algorithms to enforce cycle constraints [11, 12, 6].

4. Planar Cycle Coverings

Consider a planar embedding of the graph $\mathcal{G}$ corresponding to a binary MRF. Since we cannot directly connect the unary node X_0 to every node in the graph without losing planarity, we propose the following relaxation. For each face f of $\mathcal{G}$ add an independent copy of the unary node X_0^f and connect it to all vertices on the boundary of the face with weights θ_i^f; see Figure 1 above. Let N_i be the set of unary node copies attached to node i. We split the original unary potential θ_i across all the unary face nodes connected to i while maintaining the constraint that $\sum_{f \in N_i} \theta_i^f = \theta_i$. Using this system we have the following relaxation

$$E_{MAP} = \min_{X:X_0^f = X_0} \sum_{i>j} \theta_{ij}[X_i \neq X_j] + \sum_{i,f} \theta_i^f[X_i \neq X_0^f]$$

$$\geq \min_X \sum_{i>j} \theta_{ij}[X_i \neq X_j] + \sum_{i,f} \theta_i^f[X_i \neq X_0^f] \quad (5)$$

The inequality arises because we have dropped the constraint that all copies of X_0 take on the same value. On the other hand, since the graph corresponding to the relaxation in Equation 5 is planar, we can compute the minimum exactly. Furthermore, we have freedom to adjust the θ_i^f parameters so long as they sum up to our original parameters. This yields the variational problem

$$E_{PCC} = \max_{\theta: \sum_f \theta_i^f = \theta_i} \min_X \sum_{i>j} \theta_{ij}[X_i \neq X_j] + \sum_{i,f} \theta_i^f[X_i \neq X_0^f]$$

$$(6)$$

where $E_{MAP} \geq E_{PCC}$.

Although this planar decomposition includes duplicate copies of nodes from the original problem, it differs from standard dual decomposition in that there are not multiple independent subproblems but just a single, larger planar problem to be solved. This is analogous to our work in [15] which replaces the collection of spanning trees in TRW with a single "covering tree". We refer to this construction as a *planar cycle covering* of the original graph since the singular potentials for each face cycle are covered by some auxiliary node and in fact all other cycles also are covered in the following precise sense:

Theorem 4.1 *The lower-bound given by the planar cycle covering graph is equal to the lower-bound given by decomposition into the collection of all cycles.*

For a proof and details see the technical report [16].

4.1. Bound optimization

As with dual decomposition, the parameters may be optimized using subgradient or marginal fixed-point updates. For example, the subgradient updates for θ_i^f at a given setting of X can be easily computed by taking a gradient and enforcing the summation constraint. This yields the update rule

$$\theta_i^f = \theta_i^f + \alpha \left([X_i \neq X_0^f] - \frac{1}{|N_i|} \sum_{g \in N_i} [X_i \neq X_0^g] \right) \quad (7)$$

where $|N_i|$ is the number of auxiliary face nodes attached to X_i and α is a stepsize parameter. After each such gradient step, one must recompute the optimal setting of X which can be done efficiently using perfect matching.

The subgradient update lends itself to a simple interpretation. If X_0^f disagrees with X_i but the other neighboring copies $\{X_0^g\}$ do not, then the cost for X_0^f and X_i disagreeing is increased. On the other hand, if all the copies $\{X_0^g\}$ take on the same state then the update leaves the parameters unchanged.

5. Planar Subproblems for non-Binary MRFs

We would like to exploit the tractability of binary planar problems to lower-bound energies for planar MRFs where each node takes on one of d states. There is clearly not a one-to-one mapping from our d-state MRF to a binary planar problem in general. However, if the pairwise potentials happened to take on only two values across the $d \times d$ possible states along every edge, then we could first project down to an equivalent binary problem and then lift the solution back to the original state space.

To be precise, for a given subproblem k with nodes X_i^k, suppose we partition the state space of the original variables

X_i into two subsets $S_i^k \cup \bar{S}_i^k = \{1 \ldots d\}$. We allow for each node i in each subproblem k to have a distinct partition. We will say that the potentials θ^k are of the *planar binary type* if the graph $\mathcal{G}(\theta^k)$ is planar, $\theta_i^k = 0$ and θ_{ij}^k is of the form:

$$\theta_{ij}^k(X_i, X_j) = \begin{cases} \lambda_{ij}^k & : \quad (X_i^k \in S_i^k) \oplus (X_j^k \in S_j^k) \\ 0 & : \quad \text{otherwise} \end{cases}.$$

where $\oplus$ denotes exclusive-or. Then we can define a projected energy function of the form shown in Equation 2 with binary variables $\hat{X}_i^k$ where $\hat{X}_i^k = 1 \iff X_i^k \in S_i^k$ and edge weights λ_{ij}^k. The solution to this binary problem can be found efficiently using perfect matching.

We can now construct a lower-bound consisting of a tree-structured problem (as in TRW) along with the set of binary subproblems defined by partitions $\{S^1, S^2, \ldots\}$. We write this optimization problem as:

$$E_{MAP} \geq \max_{\{\theta^t, \theta^0, \theta^k\}} \min_{\{X^t, X^k\}} \sum_{i,t} \sum_u \theta_{i;u}^t X_{i;u}^t +$$

$$\sum_{(i,j)} \sum_{u,v} \theta_{ij;uv}^0 X_{i;u}^{t_{ij}} X_{j;v}^{t_{ji}} +$$

$$\sum_{(i,j),k} \sum_{u,v} \theta_{ij;uv}^k X_{i;u}^k X_{j;v}^k \quad (8)$$

subject to the constraints:

$$\sum_t \theta_{i;u}^t = \theta_{i;u} \qquad \sum_k \theta_{ij;uv}^k + \theta_{ij;uv}^0 = \theta_{ij;uv}$$

θ^k is of the planar binary type

By convention, index t always runs over the copies of nodes X_i^t in the tree structured problem and index k always runs over copies in the planar problems. We use θ_{ij}^0 to denote the allocation of the pairwise potentials to the tree structured subproblem. t_{ij} is the index of the copy of node X_i which is connected along edge ij in the tree.

In optimizing the bound, we are trying to find an allocation of the unary parameters among copies of the nodes in the tree and allocations of the pairwise parameters across the tree and binary planar subproblems. We use subgradient bundle trust techniques for solving this bound optimization problem. For a fixed value of the X variables, the function is linear in θ. In the optimization, one alternates between (1) solving the minimization for the X using dynamic programming to find $\{X^t\}$ and perfect matching to find $\{X^k\}$ and (2) updating θ using a gradient step.

6. Experimental Results and Discussion

We demonstrate the performance of the planar cycle cover bound on randomly generated binary grid problems, and compare against two state-of-the-art approaches: max-product linear programming (MPLP) with incrementally
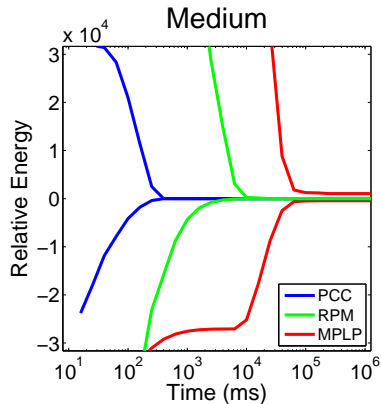
Figure 2. Average convergence behavior of lower- and upper-bounds for randomly generated 32x32 binary grid problems. We compare PCC, the planar cycle cover bound (blue) to RPM (green) and MPLP (red). Energies shown are averaged over 10 random problem instances and plotted relative to a MAP energy of 0.

added cycles [12] and reweighted perfect matching (RPM) [8]. Each problem consists a grids of size NxN with pairwise potentials drawn from a uniform distribution $\theta_{ij} \sim U(-1, 1)$. The unary potentials are generated from a uniform distribution $\theta_i \sim U(-a, a)$, where the magnitude a determines the difficulty of the problem. Figure 2 shows the upper and lower bounds found by each algorithm as a function of time, for 32×32 problem instances with $a = 0.8$. Across a range of problem instance sizes, we find that the PCC algorithm yields solutions 10-100x faster and converges to the MAP solution more frequently on hard problem instances.

In further experiments, we have demonstrated the efficacy of planar subproblems for solving non-binary MRFs. A key difficulty is in choosing a small number of such subproblems to include. We adopt a cutting-plane type strategy in which we optimize the lower-bound to convergence and then if the bound isn't tight, we use the current decoding from the tree in order to construct a partitioning of each node's state-space and add the resulting planar subproblem. This procedure guarantees that we add new cycle constraints at each iteration. In practice, for many problem instances adding a few sub-problems in this manner quickly yields a tight lower-bound. For problems with long-range cyclic dependencies, this method appears to outperform MPLP (which incrementally adds small local cycles).

We have described new variational bounds for performing inference in planar MRFs. Our bounds subsume those given by both the tree-reweighted (TRW) and outer-planar decompositions since they implicitly include every edge and cycle as a sub-problem. Unlike approaches such as MPLP which successively add cycles, we are able to get the full benefit of all cycle constraints immediately (in the binary case) or in large batches (in the d-state case). As a result we achieve fast convergence in practice.

References

[1] F. Barahona. On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical, Nuclear and General*, 15(10):3241-3253, 1982.

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

[3] A. Globerson and T. Jaakkola. Approximate inference using planar graph decomposition. In *NIPS*, pages 473–480, 2007.

[4] J. Kappes, S. Schmidt, and C. Schnoerr. MRF inference by k-fan decomposition and tight lagrangian relaxation. In *ECCV*, 2010.

[5] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Anal. Machine Intell.*, 28(10):1568–1583, 2006.

[6] N. Komodakis and N. Paragios. Beyond loose LP-relaxations: Optimizing MRFs by repairing cycles. In *ECCV*, 2008.

[7] N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *ICCV*, Rio de Janeiro, Brazil, Oct. 2007.

[8] N. Schraudolph. Polynomial-time exact inference in np-hard binary MRFs via reweighted perfect matching. In *AISTATS*, 2010.

[9] N. Schraudolph and D. Kamenetsky. Efficient exact inference in planar Ising models. Technical Report 0810.4401, Oct. 2008.

[10] W.-K. Shih, S. Wu, and Y. Kuo. Unifying maximum cut and minimum cut of a planar graph. *IEEE Transactions on Computers*, 39:694–697, 1990.

[11] D. Sontag and T. Jaakkola. New outer bounds on the marginal polytope. In *NIPS*, 2007.

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

[13] L. Torresani, V. Kolmogorov, and C. Rother. Feature correspondence via graph matching: Models and global optimization. In *ECCV*, pages 596–609, 2008.

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

[15] J. Yarkony, C. Fowlkes, and A. Ihler. Covering trees and lower-bounds on the quadratic assignment. In *CVPR*, 2010.

[16] J. Yarkony, A. Ihler, and C. Fowlkes. Planar cycle covering graphs. Technical Report arXiv:1104.1204v1, 2011.