

Beyond Trees: MRF Inference via Outer-Planar Decomposition

Dhruv Batra

Carnegie Mellon University
www.ece.cmu.edu/~dbatra

A. C. Gallagher

Eastman Kodak Company
andrew.c.gallagher@gmail.com

Devi Parikh

TTIC
dparikh@ttic.edu

Tsuhan Chen

Cornell University
tsuhan@ece.cornell.edu

Abstract

Maximum a posteriori (MAP) inference in Markov Random Fields (MRFs) is an NP-hard problem, and thus research has focussed on either finding efficiently solvable subclasses (e.g. trees), or approximate algorithms (e.g. Loopy Belief Propagation (BP) and Tree-reweighted (TRW) methods).

This paper presents a unifying perspective of these approximate techniques called “Decomposition Methods”. These are methods that decompose the given problem over a graph into tractable subproblems over subgraphs and then employ message passing over these subgraphs to merge the solutions of the subproblems into a global solution. This provides a new way of thinking about BP and TRW as successive steps in a hierarchy of decomposition methods. Using this framework, we take a principled first step towards extending this hierarchy beyond trees. We leverage a new class of graphs amenable to exact inference, called outer-planar graphs, and propose an approximate inference algorithm called Outer-Planar Decomposition (OPD).

OPD is a strict generalization of BP and TRW, and contains both of them as special cases. Our experiments show that this extension beyond trees is indeed very powerful – OPD outperforms current state-of-art inference methods on hard non-submodular synthetic problems and is competitive on real computer vision applications.

1. Introduction

A number of vision problems can be formulated as discrete labelling problems, e.g., segmentation, geometric labelling, name-face association and stereo. Markov Random Fields (MRFs) provide natural frameworks for modelling and solving these labelling problems, and thus the maximum a posteriori (MAP) inference in MRFs is of fundamental importance to vision researchers. Specifically, given a set of discrete random variables $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$ (s.t. $X_u \in L = \{1, 2, \dots, k\}$) and a graph $G = (\mathcal{V}, \mathcal{E})$ defined over these variables, the goal of MAP inference is to minimize a real-valued energy function associated with this

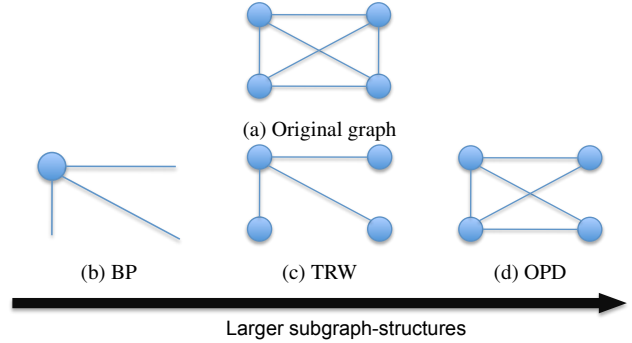


Figure 1: Decomposition Methods: BP and TRW can be seen as successive steps in a hierarchy, and OPD is a strict generalization.

graph, i.e.:

$$\mathcal{X}^* = \operatorname{argmin}_{\mathcal{X} \in \mathcal{L}} E(\mathcal{X}) \quad (1a)$$

$$= \operatorname{argmin}_{\mathcal{X} \in \mathcal{L}} \sum_{u \in \mathcal{V}} E_u(X_u) + \sum_{(u,v) \in \mathcal{E}} E_{uv}(X_u, X_v), \quad (1b)$$

where the first term $E_u(\cdot)$ is called the node or unary energy, and the second term $E_{uv}(\cdot, \cdot)$ is called the edge or pairwise energy. An equivalent way of writing this problem, which will be useful later, is in terms of boolean variables $\mathcal{Y} = \{\mathbf{Y}_u, \mathbf{Y}_{uv} \mid u \in \mathcal{V}, (u, v) \in \mathcal{E}\}$, where $\mathbf{Y}_u = \{Y_{u:s} \mid s \in L\}$, $\mathbf{Y}_{uv} = \{Y_{uv:st} \mid s, t \in L\}$.

$$\operatorname{argmin}_{\mathcal{Y} \in \mathcal{M}(G)} \sum_{u \in \mathcal{V}, s \in L} E_u(s) \cdot Y_{u:s} + \sum_{(u,v) \in \mathcal{E}, s, t \in L} E_{uv}(s, t) \cdot Y_{uv:st} \quad (2)$$

$$\mathcal{M}(G) = \left\{ \mathcal{Y} \mid \begin{array}{l} \sum_{s \in L} Y_{u:s} = 1, \quad \forall u \in \mathcal{V} \\ \sum_{t \in L} Y_{uv:st} = Y_{u:s}, \quad \forall (u, v) \in \mathcal{E}, s \in L \\ Y_{u:s} \in \{0, 1\}; Y_{uv:st} \in \{0, 1\} \end{array} \right\} \quad (3)$$

Hardness. This problem is known [32] to be NP-hard for arbitrary $G, E(\mathcal{X})$, and thus research has focussed on 1) finding efficiently solvable subclasses and 2) approximate inference algorithms for the general case.

Exact Subclasses. Trees [26] were the first known structures on which MAP could be computed efficiently via

Pearl’s Belief Propagation algorithm. For 2-class problems, exact inference can be efficiently performed via max-flow/min-cut if the energies are submodular [2, 11, 18]. More recently, Schraudolph and Kamenetsky [30] have developed on previous techniques in statistical physics [15, 22], and networks [31] to show that exact efficient inference is possible in planar Ising models with min-cut techniques even when the energies are non-submodular.

Approximate Inference. Attempts at approximate inference in general graphs have resulted in two promising techniques: 1) a straightforward application of BP to graphs with loops (Loopy BP) and 2) tree-reweighted (TRW) methods [16, 21, 35] that first decompose the problem onto trees and then obtain (approximate) global solutions via message passing schemes.

Decomposition Methods. One unifying perspective of these different approximate techniques is to think of them all as “Decomposition Methods”. These are methods that decompose the given problem over a graph into tractable subproblems over subgraphs and then employ message passing over these subgraphs to merge solutions. In Loopy BP, these subgraphs are individual nodes and their markov blankets and the message is the min-marginal. For all three of the TRW methods: TRW-T [35], TRW-S [16], TRW-DD¹ [21], these subgraphs are trees contained in the graph. In TRW-T,S these messages are min-marginals, while in TRW-DD these message are derived from MAP states on these trees. When viewed from the prism of this representation, BP and TRW can be seen as successive steps in a hierarchy of decomposition methods, and the key motivation of this paper is easy to understand – can we make these subproblems larger while still staying tractable? Fig. 1 shows an illustration.

Contributions. In this work, we leverage a class of graph structures more general than trees yet still amenable to exact inference, called outer-planar graphs, and propose a new approximate inference algorithm called Outer-Planar Decomposition (OPD). OPD follows the decomposition paradigm introduced above, *i.e.*, it first decomposes an energy function into energies over outer-planar graphs, and then use a message passing algorithm to merge these solutions. OPD is a generalization of tree-based methods and BP. In fact, for particular choices of outer-planar graphs and message passing schemes OPD reduces to each of the TRW methods (TRW-T, TRW-S, TRW-DD) and BP. Our experiments show that this extension beyond trees is indeed very powerful – OPD outperforms current state of art inference methods (TRW, QPBO and BP) on hard non-submodular synthetic problems, and is competitive on real vision applications and standard benchmarks. In addition, we introduce to the vision community certain results well-known in the

combinatorics and graph theory literature regarding outer-planar graphs and finding outerplanar subgraphs within arbitrary graphs. On a broader scope, we suggest that topological aspects of graphs are fertile grounds for improving and gaining a better understanding of inference in MRFs.

2. Relations to Previous Work

Tightening tree-decompositions. A number of works have proposed ways to improve tree-based formulations, whether working directly off the graph or via the LP relaxation of the integer programming formulation of MAP and its Lagrangian dual. As an extension of their TRW-T work, Wainwright *et al.* [35] suggested that hyper-trees could be used instead of trees to get progressively tighter bounds on the Lagrangian dual. Sontag and Jaakkola [33] suggest tightening the TRW approximation by finding a more accurate (or tighter) outer bound of the marginal polytope than the local polytope approximation of TRW methods. Komodakis and Paragios [19] work with a family of tighter relaxations called cycle-relaxations that is based on the idea of “repairing inconsistent cycles”. Sontag *et al.* [34] have proposed tightening the standard relaxation via cluster-based LP relaxations, where local consistency is achieved between cluster marginals. More recently, Komodakis and Paragios [20] have used hypergraph decompositions to get tighter bounds on the Lagrangian dual. Interestingly, while their method is general enough to allow for any hypergraph, the specific decomposition choices in their work are in fact contained in the class of outer-planar graphs. Werner [36, 37] showed that the standard TRW-LP relaxation can be improved by adding short cycles, and suggested that adding more complex tractable subproblems may help. We believe outer-planar graphs provide a good starting point. One of the goals of this paper is to evaluate the behaviour and improvement of these relaxations when using structures that are topologically more complex than trees.

It is also worth mentioning the work of Globerson and Jaakkola [8], who generalized some of the ideas of TRW-T [35] from trees to planar Ising models (which allow for the efficient computation of the partition function in closed form). At a high level, the philosophy of their approach (*i.e.* decomposition into tractable subproblems) is similar to the one presented in this paper. However, while their goal is to approximate the log-partition-function, we focus on MAP inference, and the actual techniques developed in the two papers are quite different.

Topology or parameters? Many works in vision have focussed on finding subclasses of energies that are amenable to exact efficient inference, *e.g.* boolean submodular energies [18], multi-label submodular energies [28], convex priors over a linearly ordered label set [12]. The goal of this paper is to explore the role of topological aspects of graphs, without imposing constraints on the energies. We note that

¹Komodakis *et al.* [21] refer to their method as DD-MRF, but for the sake of a common notation we call their technique TRW-DD in this paper.

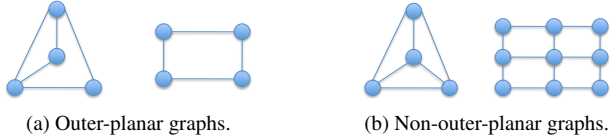


Figure 2

recent works [13, 14] have begun to explore these issues.

Organization of paper. The rest of this paper is organized as follows: §3 introduces outer-planar graphs and related theoretical properties; §4 presents our approach (OPD); §5 shows results on synthetic energies and real applications; Finally, §6 concludes with discussions.

3. Outer-Planar Graphs

Outer-Planar graphs. Outer-planarity is a notion from classical graph theory. Outer-planar graphs are a sub-class of planar graphs – they allow a planar embedding (drawing) in which all nodes lie on an external unbounded face (*i.e.* are “accessible” from the outside without crossing edges). An alternate definition is sometimes more intuitive. A graph $G = (\mathcal{V}, \mathcal{E})$ is outer-planar iff the modified graph $G_c = (\mathcal{V} \cup \{o\}, \mathcal{E} \cup \{(o, v) : v \in \mathcal{V}\})$ formed by adding an extra node o and connecting it to all nodes in G is planar. Fig. 2 shows some examples of outer-planar graphs. Outer-planar graphs are a strict generalization of trees and forests, *i.e.*, all trees and forests are outer-planar, but all outer-planar graphs are not trees/forests. The most popular graph in vision, the grid-graph, is planar but not outer-planar.

Graph Covering. The first task in any decomposition method is to find a set of subgraphs $H = \{G^i(\mathcal{V}^i, \mathcal{E}^i) : i \in \mathcal{I}\}$ that “cover” the graph G (*i.e.*, $\cup \mathcal{V}^i = \mathcal{V}$ and $\cup \mathcal{E}^i = \mathcal{E}$). In the case of TRW methods, a typical choice is a collection of spanning trees.² However, while all spanning tree subgraphs of G contain the same number of edges, all spanning outer-planar subgraphs do not. Thus, it becomes interesting to ask if we can find “densest” outer-planar subgraphs in G . Unfortunately, the problem of finding maximum outer-planar subgraphs in a general graph is NP-complete [38] but effective heuristics have been proposed [4, 27]. Another interesting question to ask is – how many such subgraphs are required to cover G ?

Outerthickness and Arboricity. The outerthickness of a graph $\Theta_o(G)$ is defined as the minimum number of outer-planar subgraphs $G^i = (\mathcal{V}^i, \mathcal{E}^i)$ of G such that the union of all the subgraph node/edge sets is equal to the node/edge set of G (*i.e.*, $\cup \mathcal{V}^i = \mathcal{V}$ and $\cup \mathcal{E}^i = \mathcal{E}$). The equivalent definition for trees is arboricity, *i.e.*, the minimum number of trees required to cover G . Arboricity and outerthickness of G are relevant for us because they denote the minimum number of subgraphs on which we would need to perform exact inference.

²Monotonic chains were used in TRW-S [16] for efficiency.

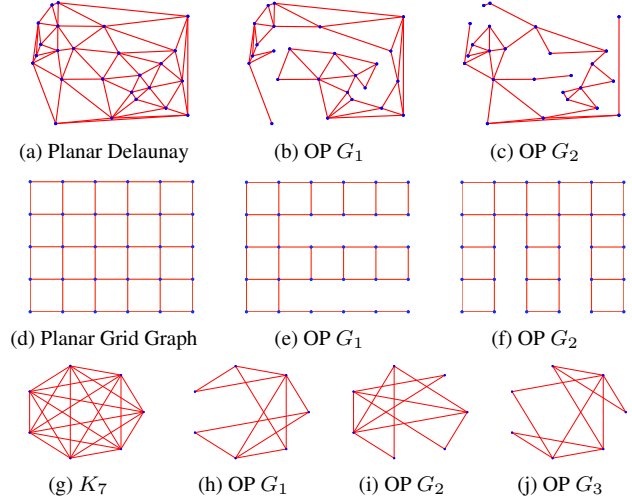


Figure 3: Example outer-planar graph decompositions.

There are several interesting results known about outerthickness. It is well known that an outer-planar graph on n vertices can have at most $2n - 3$ edges. Thus one obvious lower bound is $\Theta_o(G) \geq \lceil \frac{|\mathcal{E}|}{2n-3} \rceil$. For general graphs, even computing outerthickness is conjectured to be NP-hard [23]. However, for several classes of graphs including complete graphs and planar graphs, the results are known. For complete graphs with n vertices, the outerthickness is $\Theta_o(K_n) = \lceil \frac{n+1}{4} \rceil$, except for K_7 , where $\Theta_o(K_7) = 3$ [10].

It is worth mentioning that a long-standing conjecture of Chartrand [3] was recently proven by Gonçalves [9], which states that the outerthickness of planar graphs is two. This result is particularly powerful for vision researchers, because it means that for typical vision applications (working with planar MRFs: grid or superpixel-adjacency) we require as few as *two* outerplanar subgraphs to cover G .

To the best of our knowledge, there exists no actual implementation that is guaranteed to decompose G into $\Theta_o(G)$ number of outer-planar subgraphs, even when the theoretical result is known (*e.g.*, planar or complete graphs). Greedy heuristics are used for this purpose [4, 27]. For example, a *maximal* outer-planar subgraph G_i is created by finding a spanning tree and adding edges one at a time till no more edges can be added without violating outerplanarity (which can be easily tested [24]). Then a graph $G' = G - G_i$ is created by removing the edges of G_i from the G and finding a new maximal outerplanar graph in G' . The process is repeated until all edges of G are covered. We implemented some of these heuristics. For planar graphs, our heuristics are able to find a 2-subgraph decomposition surprisingly often (> 95%), and never found more than a 3-subgraph decomposition in our experiments. For grid graphs, we show a deterministic 2-subgraph decomposition in Fig. 3, which we call a 2-comb decomposition.

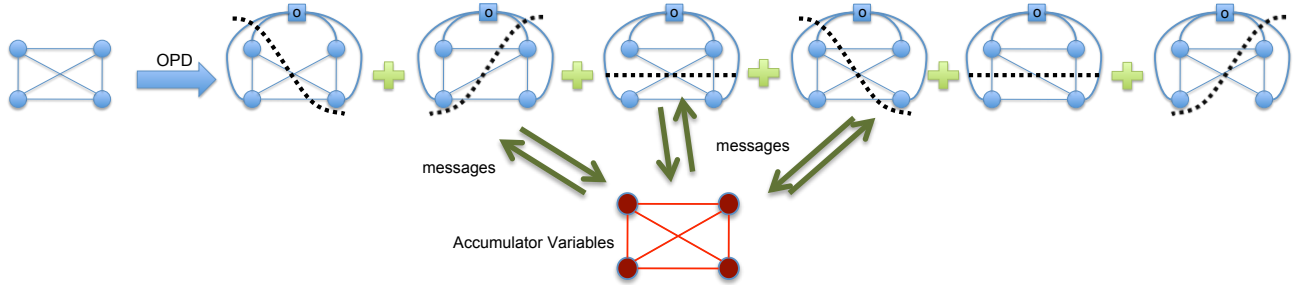


Figure 4: OPD decomposes the energy minimization problem over a general (non outer-planar) graph (K_4) into subproblems over outer-planar subgraphs. Exact inference on these outer-planar graphs is then performed. Finally a message passing scheme (§4) is used over these outer-planar subgraphs to compute (approximate) global solutions.

4. Outer-Planar Decomposition

In this section, we develop our proposed approximate inference algorithm – Outer-Planar Decomposition. OPD follows the decomposition paradigm introduced before (§1), *i.e.*, it first decomposes an energy function into energies over outer-planar graphs, performs exact inference over these outer-planar graphs, and then uses a message passing algorithm to merge these solutions.

Specifically, let $\mathcal{O} = \{O^1, \dots, O^p\}$ be the collection of outer-planar graphs that cover G . Our analysis does not require these subgraphs \mathcal{O} to be spanning, and let us associate each subgraph O^i with its own set of variables ($\mathcal{X}^i \subseteq \mathcal{X}$ or $\mathcal{Y}^i \subseteq \mathcal{Y}$), node energies θ_u^i and edge energies θ_{uv}^i . In addition, we define $\mathcal{O}(u)$, $\mathcal{O}(uv)$ to be the set of outer-planar graphs that contain node u and edge (u, v) respectively. We decompose the original energy function $E(\mathcal{X} | \theta)$ into a collection of energies $\{E(\mathcal{X}^i | \theta^i)\}$ defined over each of these outer-planar subgraphs O^i s.t.:

$$\sum_{i \in \mathcal{O}(u)} \theta_u^i = \theta_u \quad \text{and} \quad \sum_{i \in \mathcal{O}(uv)} \theta_{uv}^i = \theta_{uv} \quad (4)$$

$$\implies E(\mathcal{X} | \theta) = \sum_{i=1}^p E(\mathcal{X}^i | \theta^i) \quad (5)$$

This can be easily satisfied by distributing the node and edge energies “evenly”, *i.e.*, $\theta_u^i = \frac{\theta_u}{|\mathcal{O}(u)|}$, $\theta_{uv}^i = \frac{\theta_{uv}}{|\mathcal{O}(uv)|}$. This decomposition is pictorially shown in Fig. 4.

Exact Inference in Outer-planar MRFs. For the rest of this discussion, we assume that there exists a black-box algorithm $(\mathcal{X}^{i*}, \hat{\theta}^i) = \mathcal{B}(\theta^i, O^i)$ that takes as input MRF parameters θ^i over an outer-planar graph O^i , and returns the MAP states \mathcal{X}^{i*} and min-marginals $\hat{\theta}^i$. This black-box algorithm forms the core computational element in our message passing algorithms. Outer-planar graphs have low treewidth ($=2$), and thus the junction tree algorithm is an attractive option. More discussion regarding OPD and low treewidth models can be found in §6. For our experiments, we use the publicly available planar Ising inference engine of Schraudolph and Kamenetsky [30], which is an efficient

large-scale implementation capable of handling millions of nodes. For details, please see our extension [1] to a dynamic version of this algorithm that also efficiently computes min-marginals for outer-planar graphs. To give an idea of speed, our implementation takes just under 5 seconds to compute min-marginals on a 1000 node outer-planar graph.

Thus, we have now decomposed our original intractable energy minimization problem into energies over tractable outer-planar subgraphs. Next, we develop message passing algorithms that operate over these subgraphs.

4.1. OPD-MP: Generalization of BP

We start with the simplest message passing that generalizes the classical max-product BP algorithm and is in part motivated by work of Duchi *et al.* [5], who showed how a max-product message passing scheme could be constructed over subnetworks in a graph. This message passing scheme works in two phases. First, each subgraph i computes a min-marginal-based message ($\delta_{u:s}^{A \rightarrow i}$) that represents the belief that a node X_u^i takes state s . These messages are passed to a global accumulator \mathcal{A} . In the second phase, the accumulator sends back a message ($\delta_{u:s}^{A \rightarrow i}$) to each subgraph i , which represents the sum of beliefs of all other subgraphs. For ease of notation, we use vector forms of these messages ($\delta_u^{A \rightarrow i} = \{\delta_{u:1}^{A \rightarrow i} \dots \delta_{u:k}^{A \rightarrow i}\}$). The procedure is shown in Alg. 1. The intuition behind these messages is simple – after the initial decomposition of energies, different subgraphs might assign different states to the same variable X_u . The goal then is to force them to agree on this assignment, by incorporating the accumulated beliefs. It is not difficult to see that OPD-MP reduces to regular BP for a particular choice of subgraphs (a collection of star-graphs centered at each node in a factor-graph representation).

4.2. OPD-DD: Generalization of TRW-DD

OPD-DD uses the message passing scheme introduced by Komodakis *et al.* [20, 21], and is based upon dual decomposition ideas from the optimization literature. Conceptually, this is a different scheme than the previous one

```

Initialize  $\delta_v^{A \rightarrow k} = 0, \forall v \in \mathcal{V}^k, k \in \mathcal{O}(v)$ 
Initialize  $\lambda_{uv}^k = \theta_{uv}^k, \forall (u, v) \in \mathcal{E}^k, k \in \mathcal{O}(uv)$ 
while Stopping criteria not met do
  1. Solving Subproblems: for  $i = 1$  to  $p$  do
    for  $u \in \mathcal{V}^i$  do
       $\lambda_v^i = \theta_v^i + \delta_v^{A \rightarrow k}, \forall v \in \mathcal{V}^i \setminus u$  (6)
       $\lambda_u^i = \theta_u^i$  (7)
       $\delta_{u:s}^{i \rightarrow A} = \min_{\mathcal{X}^i \in \mathcal{L}; X_u^i = s} E(\mathcal{X}^i | \lambda^i)$  (8)
    end
  end
  2. Message Passing:
   $\delta_u^{A \rightarrow i} = \sum_{j \neq i} \delta_u^{j \rightarrow A}, \forall i \in \{1, \dots, k\}$  (9)
end

```

Algorithm 1: OPD-MP.

because it only uses the MAP states from the subgraphs to construct messages, and has interesting resource allocation interpretations [20, 21]. The procedure is shown in Alg. 2.

It can be easily shown that OPD-DD corresponds to a projected subgradient descent algorithm for solving the following Lagrangian dual problem $\mathcal{F}(\{\lambda^i\})$:

$$\max_{\{\lambda^i\} \in \Lambda} \mathcal{F}(\{\lambda^i\}) = \sum_{i=1}^p \mathcal{F}^i(\lambda^i) \quad (10)$$

$$\mathcal{F}^i(\lambda^i) = \min_{\mathcal{X}^i} E(\mathcal{X}^i | (\theta^i + \lambda^i)) \quad \text{where,} \quad (11)$$

$$\Lambda = \left\{ \{\lambda^i\} : \sum_{i \in \mathcal{O}(u)} \lambda_u^i = 0, \sum_{i \in \mathcal{O}(uv)} \lambda_{uv}^i = 0 \right\} \quad (12)$$

The above problem can be seen as a generalization of TRW-DD [21]. In fact, since trees are a subset of outer-planar graphs, OPD-DD reduces to TRW-DD for a particular decomposition choice (when all O^i are chosen as trees). In addition, we state a simple result that can be derived from the work of Komodakis *et al.* [20]: When the multipliers α_t satisfy $\lim_{t \rightarrow \infty} \alpha_t = 0, \sum_t \alpha_t = \infty$, OPD-DD converges to the global optimum of the Lagrangian dual. Also, at each iteration, $\mathcal{F}(\{\lambda^i\})$ provides a valid lower bound on the energy minimization problem.

5. Experiments

We now present experiments to compare the performance of OPD with current state of art methods. We tested OPD on the following applications: gender-face assignment, multiclass object labelling and optical flow. In addition, we also present results on synthetic energies, where we can control various parameters (*e.g.*, the degree of non-submodularity of energies) to see how OPD behaves with changes in these parameters. We compare OPD with TRW-

```

for ( $t = 1$  to  $\infty$ ; Until stopping criteria met) do
  1. Reparameterization:
  for  $i = 1$  to  $p$  do
     $(\mathcal{Y}^{i*}, \tilde{\theta}^i) = \mathcal{B}(\theta^i, O_i)$  (13)
  end
  2. Averaging:
  for  $i = 1$  to  $p$  do
     $\lambda_u^i = \alpha_t \left( \mathbf{Y}_u^* - \frac{\sum_{i \in \mathcal{O}(u)} \mathbf{Y}_u^*}{|\mathcal{O}(u)|} \right)$  (14a)
     $\lambda_{uv}^i = \alpha_t \left( \mathbf{Y}_{uv}^* - \frac{\sum_{i \in \mathcal{O}(uv)} \mathbf{Y}_{uv}^*}{|\mathcal{O}(uv)|} \right)$  (14b)
     $\theta_u^i += \lambda_u^i; \quad \theta_{uv}^i += \lambda_{uv}^i$  (14c)
  end
end

```

Algorithm 2: OPD-DD.

S [16], BP and QPBO [29], using implementations provided by authors. In case of non-submodular problems, when QPBO leaves nodes unlabelled, we use the following heuristic – we set labelled nodes to the states returned by QPBO, and create a smaller energy minimization problem on the unlabelled nodes. We then solve this smaller problem with TRW-S, and refer to this heuristic as QPBO+h.

Synthetic Energies. Following the setup of Kolmogorov [16], we create artificial energy functions by randomly sampling from Gaussians: $\theta_{u:0}, \theta_{u:1} \sim \mathcal{N}(0, 1)$. Pairwise energies were set as follows: $\theta_{uv:00} = \theta_{uv:11} = 0$; $\theta_{uv:01}, \theta_{uv:10} \sim \mathcal{N}(0, \sigma^2)$ and $\sigma = \{\frac{1}{d}, \frac{3}{d}, \frac{5}{d}\}$, where d is the max-degree in G . We note that these edge-terms are mixed potentials, *i.e.*, some of them are submodular, while others are supermodular. Increasing the value of σ increases the effect of interaction/edge terms, and generally makes the problem harder to solve. For example, a typical effect is that QPBO labels fewer and fewer nodes as σ increases (at $\sigma = 5/d$, QPBO always left all nodes unlabelled). For all our experiments, we generated 100 such energies, and report mean results.

Fig. 5a shows the performance of various methods on a 4-node fully connected graph, which is the smallest non-outer-planar graph. We use the 6-subgraph decomposition shown in Fig. 4. First we note that when σ is small ($1/d$), all methods are able to solve the problem, and the gap between the lower-bounds and energies becomes very small. However, with larger σ the behaviour is different. QPBO labels fewer nodes (for $\sigma = 5/d$, QPBO always left all nodes unlabelled). We can see that the lower-bound of OPD-DD is tighter than TRW-S and the energies are lower than others.

Figs. 5b, 5c shows results on a 50-node fully connected graph (K_{50}). We use the heuristic described in §3 to find the collection of outer-planar subgraphs that cover K_{50} . Our method typically needed 15 spanning outer-planar sub-

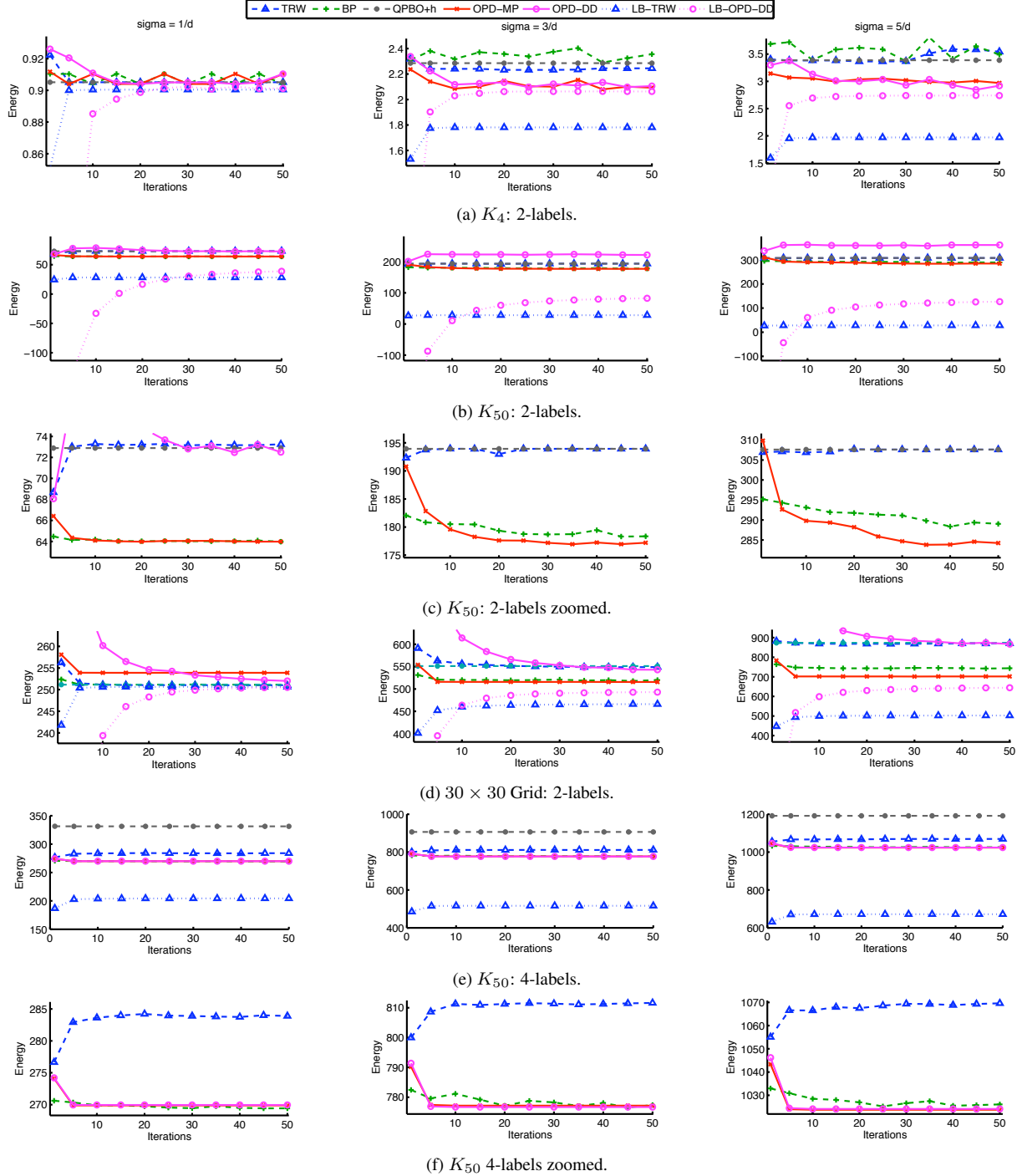


Figure 5: Synthetic Energies: Each row corresponds to particular graph structure. Columns corresponds to different (increasing) values of σ (see §5 for details about energy construction). As σ increases the performance gap between OPD and baselines becomes more evident.

graphs, which is only slightly higher than the outerthickness $\Theta_o(K_{50}) = \lceil \frac{50+1}{4} \rceil = 13$. As in the previous experiment, OPD-MP performs the best and the lower-bound of OPD-DD is the tightest, even though the primal solution of OPD-DD does not make much progress. This suggests that the Lagrangian relaxation of OPD-DD may not be tight in this

case. We note this corroborates with a similar observation made by Kolmogorov [16].

Fig. 5d shows results on a 30x30 grid. We use the 2-comb decomposition shown in Fig. 3 to get 2 outer-planar graphs. Again, OPD-MP performs the best and the OPD-DD lower-bound is the tightest. In addition, OPD-DD pri-

mal does not stall, which suggests that grids may be better for the OPD-DD Lagrangian relaxation.

Fig. 5e shows the results for a 4-label problem on K_{50} . In this case, multi-label energies were randomly sampled in a similar manner as before. For this experiment, QPBO and OPD were used within an alpha expansion framework. The multi-label energies always violated some triangle inequalities, and we did not perform any truncation. We can see that AE-OPD-MP/DD both perform very similar and do better than TRW-S and BP. It is important to mention that the iteration number (x-axis) for AE-QPBP/OPD refers to iterations of alpha-expansion and is not directly comparable with TRW/BP iterations.

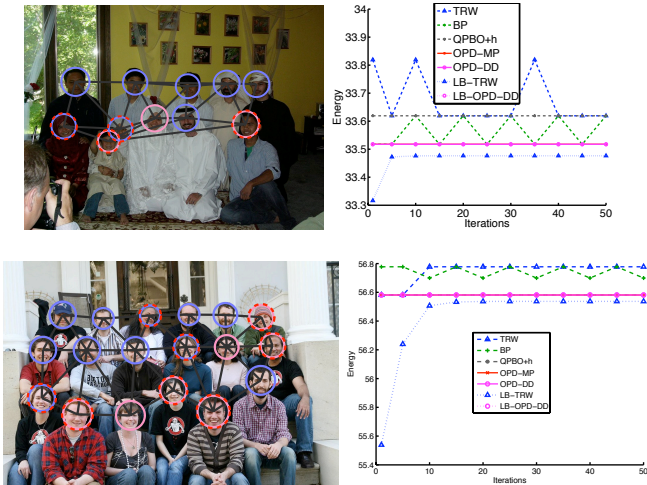


Figure 6: Gender-Face Assignment: The MAP solution over the defined graph is shown using pink for female and blue for male. Red dashed circles show incorrect gender classification.

We now report results on real applications. In all cases OPD performed competitive or better than baselines. **Gender-Face Assignment.** In this application, the goal is to assign a gender to all faces found in a group photograph. A graph is formed by first detecting faces. Each face is considered a node and edges are found with Delaunay triangulation. The node energy at a face corresponds to the negative log-likelihood that this face is assigned to a particular gender, using a gender classifier based on facial appearance. Edge features capturing the relative sizes and positions of the two faces corresponding to an edge are extracted. The frequency of gender pair assignments for 200-nearest neighbours in a labelled dataset [6], is used to create the edge energy. More details about the dataset, features used, appearance classifiers and pairwise descriptors can be found in Gallagher and Chen [7]. Fig. 6 shows the energy plots, and the gender assignments of OPD-MP. We note that OPD-DD (and MP) finds the global solution (lower-bound of OPD-DD = energy) in the first iteration for these instances.

Multiclass Object Labelling. In this application, the task is to label each segment in an image (in the MSRC object

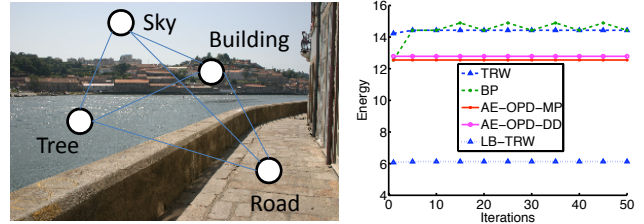


Figure 7: Object Labelling (MSRC).

class recognition dataset) with one of 21 object categories such as cow, grass, etc. A fully connected graph is constructed over these segments. The node energies are computed as the negative log of the likelihood of that segment being assigned to one of the 21 classes, using a classifier trained on texture and color features. The edge energies are the negative log of the class co-occurrence matrix from the training data. More details about the features and appearance classifiers can be found in Parikh *et al.* [25]. Fig. 7 shows the results.

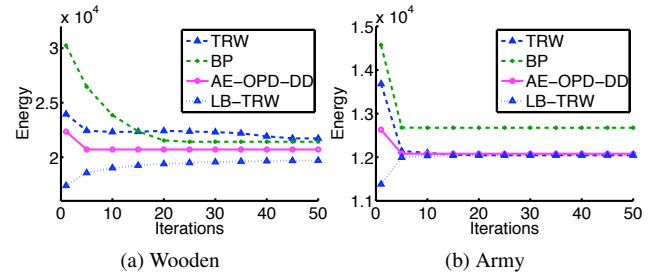


Figure 8: Optical Flow: Energy and lower bounds.

Optical Flow. We also tested OPD on a standard Middlebury optical flow benchmark. We closely followed the energy construction of Komodakis and Paragios [21]. The graph used was a 4-connected grid graph over pixels. The labels correspond to 2D motion vectors within a window. Unary energies are given by absolute intensity difference between a pixel and its corresponding location (under a flow labelling) in the next frame. Pairwise energies were a truncated weighted squared euclidean distance between the two labellings at the nodes. Fig. 8 shows the results.

6. Discussions and Future Work

To summarize, this paper presents a unifying perspective of different approximate MAP inference techniques – called “Decomposition Methods”, *i.e.* methods that decompose the given problem over a graph into tractable sub-problems over subgraphs and then employ message passing over these subgraphs to merge solutions. Using this framework, we leverage a new class of graphs amenable to exact inference (called outer-planar graphs), and propose a new approximate inference algorithm (OPD). OPD is a generalization of tree-based methods and contains TRW, BP as special cases. Our experiments show that this extension be-

yond trees is indeed very powerful – OPD outperforms current state of art inference methods on hard non-submodular synthetic problems, and is competitive on real vision applications. On a broader scope, we suggest that topological aspects of graphs are fertile grounds for improving and gaining a better understanding of MAP inference.

Generalized k-treewidth decomposition. As we mentioned earlier, outer-planar graphs are low treewidth structures. All outer-planar graphs have treewidth (TW) 2. However, the class of TW-2 structures (called series-parallel graphs) is strictly larger than outer-planar graphs (meaning that there exist TW-2 structures that are not outer-planar). It is, of course, an interesting direction to extend the ideas presented in this paper to TW-2, or more generally TW-k structures. Naturally, choosing more complex subproblems should lead to tighter approximations but inference on each of these subproblems is exponential in k, *i.e.* $O(nL^k)$. Thus each higher order decomposition comes at an additional computational cost of $O(L)$ as well as the cost of finding the higher order decomposition (however fewer decompositions will typically be required to cover the graph). So it might be a good idea to find a sweet spot required for our applications.

Acknowledgements: We thank Nic Schraudolph and Vladimir Kolmogorov for helpful discussions, for explaining us their work, and for making *isinf* [30] and BlossomV [17] respectively publicly available. We thank Pedro Felzenszwalb for pointing out the connections between outer-planar graphs and k-trees, and Carlos Guestrin and Ramin Zabih for helpful discussions.

References

- [1] D. Batra and T. Chen. Dynamic planar-cuts: Efficient computation of min-marginals for outer-planar models. In *NIPS Workshop on Discrete Optimization in Machine Learning*, 2009.
- [2] E. Boros and P. L. Hammer. Pseudo-boolean optimization. *Discrete Appl. Math.*, 123(1-3):155–225, 2002.
- [3] G. Chartrand, D. Geller, and S. Hedetniemi. Graphs with forbidden subgraphs. *J. Combinatorial Theory*, 1971.
- [4] R. Cimikowski. On heuristics for determining the thickness of a graph. *Information Sciences*, 1995.
- [5] J. Duchi, D. Tarlow, G. Elidan, and D. Koller. Using combinatorial optimization within max-product belief propagation. In *NIPS*, 2007.
- [6] A. Gallagher and T. Chen. Estimating age, gender and identity using first name priors. In *CVPR*, 2008.
- [7] A. Gallagher and T. Chen. Understanding images of groups of people. In *CVPR*, 2009.
- [8] A. Globerson and T. Jaakkola. Approximate inference using planar graph decomposition. In *NIPS*, 2006.
- [9] D. Gonçalves. Edge partition of planar graphs into two outerplanar graphs. In *STOC*, pages 504–512. ACM, 2005.
- [10] R. Guy and R. Nowakowski. The outerthickness and outercoarseness of graphs i. the complete graph and the n-cube. *Topics in Combinatorics and Graph Theory: Essays in Honour of G. Ringel*, 1990.
- [11] P. Hammer. Some network flow problems solved with pseudo-boolean programming. *Operations Research*, 13:388–399, 1965.
- [12] H. Ishikawa. Exact optimization for markov random fields with convex priors. *PAMI*, 25(10):1333–1336, 2003.
- [13] K. Jung, P. Kohli, and D. Shah. Local rules for global map: When do they work? In *NIPS*, 2009.
- [14] K. Jung and D. Shah. Local algorithms for approximate inference in minor-excluded graphs. In *NIPS*, 2007.
- [15] P. W. Kasteleyn. Dimer statistics and phase transitions. *Journ. of Math. Phys.*, 4(2):287–293, 1963.
- [16] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.
- [17] V. Kolmogorov. Blossom V: a new implementation of a minimum cost perfect matching algorithm. *Mathematical Programming Computation*, 1(1):43–67, 2009.
- [18] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.
- [19] N. Komodakis and N. Paragios. Beyond loose lp-relaxations: Optimizing mrfs by repairing cycles. In *ECCV*, 2008.
- [20] N. Komodakis and N. Paragios. Beyond pairwise energies: Efficient optimization for higher-order mrfs. In *CVPR*, 2009.
- [21] N. Komodakis, N. Paragios, and G. Tziritas. Mrf optimization via dual decomposition: Message-passing revisited. In *ICCV*, 2007.
- [22] F. Liers and G. Pardella. A simple max-cut algorithm for planar graphs. Technical report, Combinatorial Optimization in Physics (COPhy), 2008.
- [23] E. Mäkinen and T. Poranen. An annotated bibliography on the thickness, outerthickness, and arboricity of a graph. Technical report, University of Tampere, 2009.
- [24] S. L. Mitchell. Linear algorithms to recognize outerplanar and maximal outerplanar graphs. *Information Processing Letters*, 9(5):229–232, 1979.
- [25] D. Parikh, C. Zitnick, and T. Chen. From appearance to context-based recognition: Dense labeling in small images. In *CVPR*, pages 1–8, 2008.
- [26] J. Pearl. Reverend bayes on inference engines: A distributed hierarchical approach. In *AAAI*, 1982.
- [27] T. Poranen. *Approximation algorithms for some topological invariants of graphs*. PhD thesis, University of Tampere, 2004.
- [28] S. Ramalingam, P. Kohli, K. Alahari, and P. H. S. Torr. Exact inference in multi-label crfs with higher order cliques. *CVPR*, 2008.
- [29] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary mrfs via extended roof duality. In *CVPR*, June 2007.
- [30] N. N. Schraudolph and D. Kamenetsky. Efficient exact inference in planar ising models. In *NIPS*, 2008.
- [31] W.-K. Shih, S. Wu, and Y. S. Kuo. Unifying maximum cut and minimum cut of a planar graph. *IEEE Trans. Comput.*, 39(5):694–697, 1990.
- [32] S. E. Shimony. Finding maps for belief networks is np-hard. *Artificial Intelligence*, 68(2):399–410, August 1994.
- [33] D. Sontag and T. Jaakkola. New outer bounds on the marginal polytope. In *NIPS*, 2007.
- [34] D. Sontag, T. Meltzer, A. Globerson, T. Jaakkola, and Y. Weiss. Tightening lp relaxations for map using message passing. In *UAI*, 2008.
- [35] M. Wainwright, T. Jaakkola, and A. Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches. *IEEE Trans. Inf. Th.*, 51(11):3697–3717, 2005.
- [36] T. Werner. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (map-mrf). In *CVPR*, 2008.
- [37] T. Werner. Revisiting the decomposition approach to inference in exponential families and graphical models. Technical Report Research report CTU-CMP-2009-06, Center for Machine Perception, Czech Technical University, 2009.
- [38] M. Yannakakis. Node- and edge-deletion np-complete problems. In *Proc. ACM Symposium on Theory of Computing*, 1978.