

# Inference for Order Reduction in Markov Random Fields

Andrew C. Gallagher

Eastman Kodak Company  
andrew.gallagher@kodak.com

Dhruv Batra

Toyota Technological Institute at Chicago  
ttic.uchicago.edu/~dbatra

Devi Parikh

Toyota Technological Institute at Chicago  
dparikh@ttic.edu

## Abstract

This paper presents an algorithm for order reduction of factors in High-Order Markov Random Fields (HOMRFs). Standard techniques for transforming arbitrary high-order factors into pairwise ones have been known for a long time.

In this work, we take a fresh look at this problem with the following motivation: It is important to keep in mind that order reduction is followed by an inference procedure on the order-reduced MRF. Since there are many possible ways of performing order reduction, a technique that generates “easier” pairwise inference problems is a better reduction.

With this motivation in mind, we introduce a new algorithm called Order Reduction Inference (ORI) that searches over a space of order reduction methods to minimize the difficulty of the resultant pairwise inference problem. We set up this search problem as an energy minimization problem. We show that application of ORI for order reduction outperforms known order reduction techniques both in simulated problems and in real-world vision applications.

## 1. Introduction

The introduction of sophisticated discrete optimization tools for inference in Markov Random Fields (MRFs) over the last two decades has brought about a paradigm shift in the way we think about modern computer vision problems. It is now a routine process to write an explicit energy function, understand the Bayesian priors it incorporates, and then depending on its properties, perform exact or approximate inference. More recently, a second wave of success can be attributed to the incorporation of higher-order terms that have the ability to encode significantly more sophisticated priors and structural dependencies between variables, *e.g.*, second-order smoothness priors in stereo [24], priors on affine motions in optical-flow [4], priors on natural image statistics for de-noising [18], robust smoothness priors for object labelling [9], co-occurrence priors for binary texture restoration [3], co-occurrence priors for object category segmentation [13], connectivity [15] and bounding-box [14] priors for image segmentation.

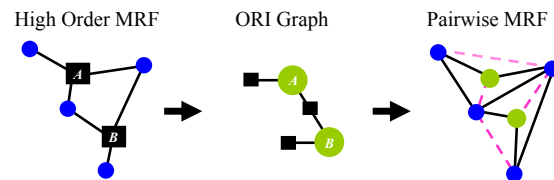


Figure 1. Existing works for order reduction in high-order MRFs apply the same order reduction to each factor. However, multiple factors in the graph can share pairwise edges in the order-reduced pairwise MRF. We propose Order Reduction Inference (ORI) to select the order-reduction method at each factor. To do this, a secondary graph, called the ORI graph, is constructed (middle) to select each factor’s order-reduction method. The resulting pairwise MRF (right) is produced from the selected method for each factor.

The classical approach for inference in Higher-Order MRFs (HOMRFs) is to first convert the higher-order terms into quadratic ones by introducing auxiliary variables into the problem, and then using standard tools (QPBO, TRW, BP) to minimize the quadratic energy functions.

This paper focuses on the first step, *i.e.*, order reduction in HOMRFs. Standard techniques for transforming arbitrary high-order factors into pairwise ones have been known for a long time — see for example, Appendix E.3 in Wainwright and Jordan [23] and Rosenberg [17]. However, a number of works have begun to revisit order reduction in HOMRFs. For instance, Ishikawa [7] recently proposed a new order-reduction technique that was justified to be better than Variable Substitution (VS) [17] because it produces fewer non-submodular terms than VS, which makes the subsequent pairwise inference easier.

**Goal.** We believe that the point implicitly made by Ishikawa’s order-reduction technique [7] is a critical one. It is important to keep in mind that order reduction will be followed by an inference procedure on the order-reduced MRF. Since there are many possible ways to perform order reduction, a technique that generates “easier” pairwise inference problems is a better reduction. In this work, we make this implicit point an explicit goal. Specifically, our goal is to search over a space of order-reduction methods to minimize the “difficulty” of the resultant pairwise inference problem. We formalize what we mean by easy and difficult

inference problems in Sec. 3.1.

**Technique.** We approach order reduction as an optimization problem. Specifically, we propose a family of new order-reduction functions that allow each higher-order factor to choose its own particular reduction, as illustrated in Figure 1. This is in contrast with previous works where *all* factors are reduced with the same reduction. We set up this selection problem as an energy minimization problem, and call our proposed algorithm Order Reduction Inference (ORI). While it may seem counter-intuitive to solve an inference problem (ORI) as an intermediate step to solve another inference problem (HOMRF), we show that the ORI problem is typically significantly smaller than HOMRF and need not be solved optimally to obtain gains with HOMRF inference.

Our experiments show that ORI’s improvement over typical order reduction (*i.e.*, applying the same order reduction method to all higher-order factors) is indeed significant, and observed across a broad range of approximate inference algorithms. As a secondary contribution of our experiments, we present an empirical comparison of various order-reduction techniques in the literature.

## 2. Higher-Order MRFs and Order Reduction

**Notation:** We first describe the notation that will be used throughout the paper.

**HOMRF.** Let  $[n]$  denote the set  $\{1, 2, \dots, n\}$ . We consider a set of Boolean random variables  $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ ,  $x_i \in \{0, 1\}$ . We emphasize that this is without loss of generality because any multi-label energy function can be converted into a boolean one via standard techniques [16, 21]. For any set  $A \subseteq [n]$ , let  $x_A$  be shorthand for  $\{x_i \mid i \in A\}$ . The input to all order-reduction algorithms is a HOMRF hypergraph defined over these variables  $\mathcal{H} = (\mathcal{V}_{\mathcal{H}}, \mathcal{E}_{\mathcal{H}})$ . We will refer to the vertices of this hypergraph as **variables**  $\mathcal{V}_{\mathcal{H}} = [n]$ , and its hyperedges as **factors**  $\mathcal{E}_{\mathcal{H}} \subseteq \mathbb{P}([n])$ , where  $\mathbb{P}([n]) = \{A \mid A \subseteq [n]\}$  is the powerset on  $n$  elements.

The goal of maximum *a posteriori* (MAP) inference in HOMRFs is to minimize a real-valued energy function:  $E(\mathcal{X}) = \sum_{A \in \mathcal{E}_{\mathcal{H}}} E_A(x_A)$ .

**ORI Graph.** Our order-reduction algorithm, ORI, will construct an intermediate graph that we call ORI-graph  $\mathcal{G} = (\mathcal{V}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}})$ . We refer to the vertices of  $\mathcal{G}$  as **ORI-nodes**, and the edges of  $\mathcal{G}$  as **ORI-edges**. Each ORI-node  $v$  represents a factor in  $\mathcal{H}$  and is associated with a decision variable  $o_v \in \mathcal{L}$  that decides which *kind* of order reduction, out of a family of possible reductions  $\mathcal{L}$ , is performed at ORI-node  $v$ . As we explain in Sec. 3.3, we will construct an energy function  $E_{\mathcal{G}}(\mathcal{O})$  indicating the quality of the order reduction  $\mathcal{O}$ .

**Pairwise-MRF.** The result of order reduction is a Pairwise MRF (PMRF)  $\mathcal{P} = (\mathcal{V}_{\mathcal{P}}, \mathcal{E}_{\mathcal{P}})$ , defined over Boolean random variables  $\mathcal{Y} = \{y_1, \dots, y_m\}$ ,  $y_i \in \{0, 1\}$ . We refer to the vertices of this graph as **nodes** and edges of this graph as **edges**. There are  $m$  nodes in this graph ( $m \geq n$ ), with the first  $n$  nodes corresponding to the variables in the HOMRF. The remaining  $m - n$  nodes are auxiliary variables produced by the order-reduction process. The energy function associated with  $\mathcal{P}$  by design does not contain any terms higher than second-order, *i.e.*,  $E_{\mathcal{P}}(\mathcal{Y}) = \sum_{A \in \mathcal{V}_{\mathcal{P}} \cup \mathcal{E}_{\mathcal{P}}} p_A \prod_{i \in A} y_i$ . Approximate inference techniques like QPBO/BP are used to find a labeling on the  $m$  nodes of the pairwise MRF, and these labels are propagated to the  $n$  variables of the HOMRF. It will be important to remember that a term  $p_{ij}y_iy_j$  is submodular if  $p_{ij} \leq 0$ .

### 2.1. Order Reductions in Literature

It is well known [2] that any energy function on Boolean variables can be uniquely represented as a *multilinear polynomial*, *i.e.*,  $E(\mathcal{X}) = \sum_{A \in \mathcal{E}_{\mathcal{H}}} E_A(x_A) = \sum_{A \in \mathcal{E}_{\mathcal{H}}} h_A \prod_{i \in A} x_i$ , where  $h_A \in \mathbb{R}$  is a scalar multiplier.

Order-reduction methods operate by expressing each high-order term,  $h_A \prod_{i \in A} x_i$  (with  $|A| > 2$ ), as an expression with only pairwise interactions. This is made possible by introducing auxiliary variables. Moreover, the new order-reduced polynomial is guaranteed (by construction) to have the same minimum as the original high-order energy, so that minimizing this order-reduced energy leads to the same solution as the higher-order energy.

We briefly review some known reductions [2, 6, 12, 19] to understand the functions they use. In the following,  $x, y, z$  are three Boolean variables.

#### 2.1.1 Ishikawa’s Symmetric Reduction (ISH)

Ishikawa [6] proposed a generalized formula to reduce any high-order term. Specifically, for a term  $xyz$  containing three Boolean variables, an auxiliary variable  $w$  is introduced in the following reduction:

$$xyz = \min_w (-wx - wy - wz + w + \color{red}{xy + yz + xz}) \quad (1)$$

$$-xyz = \min_w (2w - wx - wy - wz) \quad (2)$$

The validity of these expressions may be verified via a truth-table. It is important to note that while the term  $-xyz$  can be reduced without introducing any non-submodular terms, for  $xyz$ , three non-submodular terms always generated (shown in red).

#### 2.1.2 Reduction by Substitution (VSUB)

In this classic reduction [17], a pair of variables  $xy$  is substituted by auxiliary variable  $w$ . The reduction relies on a

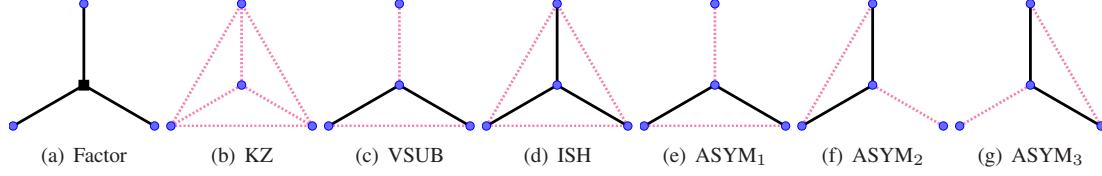


Figure 2. An illustration of reducing a triplet factor  $xyz$ . In (a), the term is shown as a factor graph. In the remaining plots, solid lines indicate submodular pairwise terms, and dashed pink lines show non-submodular terms. The reductions shown are (b) minimum selection (KZ), (c) variable substitution (VSUB), (d) Ishikawa’s symmetric reduction (ISH), (e-g) the family of asymmetric reductions introduced in this paper. Note that the selection of a reduction method implies that specific edges will have a non-submodular or a submodular contribution. The ORI algorithm exploits this property to produce a pairwise MRF that is “easy” for approximate inference algorithms.

penalty function  $D(x, y, w)$  that induces a cost only when  $w$  is not equal to  $xy$ :

$$D(x, y, w) = -2xw - 2yw + 3w + xy \quad (3)$$

One can verify via truth-tables that  $D(x, y, w) > 0$  iff  $w \neq xy$ , thus  $D(x, y, w)$  may be thought of as a disagreement indicator function. We can now replace every instance of  $xy$  by  $w$  and include this penalty function. Thus, for a large enough  $M$ :

$$\min_{x,y,z} xyz = \min_{x,y,z,w} zw + MD(x, y, w) \quad (4)$$

As noted by Ishikawa [6], this reduction is guaranteed to produce two non-submodular terms, and because  $M$  needs to be large, at least one of them ( $xy$ ) has a large magnitude coefficient.

### 2.1.3 Reduction by Minimum Selection (KZ)

Kolmogorov and Zabih [12] proposed the minimum selection reduction. For second-order terms with a positive coefficient, six non-submodular terms are produced.

$$xyz = \min_w -w - x - y - z + \quad (5)$$

$$w(x + y + z) + xy + xz + yz \quad (6)$$

$$-xyz = \min_w 2w - wx - wy - wz \quad (7)$$

## 2.2. An Asymmetric Reduction (ASM)

We now introduce a novel family of asymmetric reductions for factors of three variables:

$$xyz = \min_w w - yw - zw + xw + yz \quad (8)$$

Eq. (8) is symmetric in  $y$  and  $z$ , but not in  $x$ . This allows us to write two more asymmetric reductions:

$$xyz = \min_w w - xw - zw + yw + xz \quad (9)$$

$$xyz = \min_w w - xw - yw + zw + xy \quad (10)$$

This family of reductions ( $ASYM_1$ ,  $ASYM_2$ , and  $ASYM_3$ ) produces two non-submodular edges when reducing  $xyz$ . Consequently, by selecting either (8), (9), or (10), one can choose *which* edge on the original variables to make non-submodular. For reducing  $-xyz$ , the same reduction as in (2) is used, which does not produce any non-submodular terms. A visual representation of the order reductions is shown in Figure 2.

## 3. Inference for Order Reduction

We now describe our proposed Order Reduction Inference (ORI) algorithm. At this point, we have reviewed a number of known order-reduction methods to simplify high-order terms, and introduced a family of asymmetric reductions for triplet terms.

One key question to ask is: are some order-reduction methods superior to others? The experimental analysis of Ishikawa [6] certainly seems to suggest so — for instance, ISH resulted in pairwise energies where QPBO was able to extract partial optimality for a significantly larger set of nodes than VSUB. Is it possible to explicitly minimize the difficulty of the pairwise inference problem? Further, it is natural to ask if there is, in fact, not a single best order reduction, but rather a combination of order-reduction methods that performs well?

ORI attempts to address both of these issues by giving each high-order factor in HOMRF a choice of order reductions, *i.e.*,  $o_i \in \{KZ, VSUB, ISH, ASYM_1, ASYM_2, ASYM_3\}$ . This in turn sets up a combinatorial optimization over  $6^{|\mathcal{V}_\sigma|}$  possible choices.

### 3.1. What Makes an Inference Problem Easy?

In order to set up the ORI search problem, we need to define an objective function  $E_G(\mathcal{O})$ . Ideally, we would like ORI to choose reductions that produce “easy” pairwise energies, so that our approximate inference algorithms (QPBO / TRW / BP) perform well. While we can describe necessary conditions for an exact inference to be applicable (*e.g.*, submodularity for graph-cuts, low-treewidth graphs for junction tree), it is not a trivial problem to characterize what makes an inference problem easy for a particular *ap-*

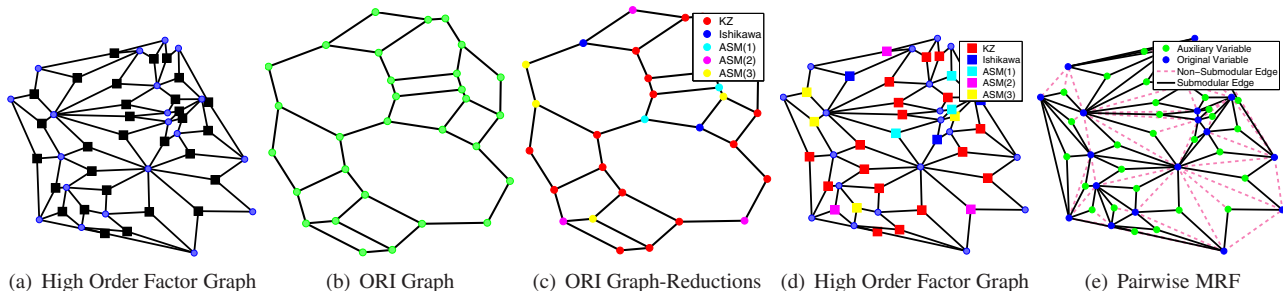


Figure 3. In the Order Reduction Inference algorithm, a global energy function is minimized to determine which order reduction to use for each high-order factor of a HOMRF. (a) The original high-order MRF (HOMRF), with black squares to indicate factors, and blue circles to indicate variables. (b) The ORI graph. Each ORI-node in the ORI graph corresponds to a factor in the HOMRF. Each ORI-edge indicates that the corresponding factors in the HOMRF share a pair of variables, whose parameters in the final PMRF are influenced by the choice of order-reduction methods at each factor. After defining the energies of the ORI graph, (c) inference on the ORI graph indicates which reduction to use for each factor of the HOMRF (shown by ORI-node color). (d) The original HOMRF, now with each factor indicating its order-reduction method. (e) The final PMRF. (Best viewed in electronic version.)

*proximate* inference algorithm. However, as several authors have noted [8, 11], we do have empirical evidence to suggest what makes an inference problem difficult — large label spaces, densely connected graphs, and non-submodular or repulsive energies. For example, Rother *et al.* [20] found that QPBO performs significantly worse than simulated annealing for the application of deconvolution, which contains densely connected graphs and repulsive potentials. Kolmogorov [10] found that TRW-S perform poorly in the presence of mixed potentials (some edges attractive, some edges repulsive). Batra *et al.* [1] found that the gap between the global minimum and the solution found by approximate inference generally increases as the weight of non-submodular edges increases.

With this intuition in mind, we explore two different heuristics for  $E_{\mathcal{G}}(\mathcal{O})$ . In the first, we try to minimize the *number* of edges in the order-reduced graph with non-submodular terms, and in the second, we try to minimize the sum of the *weights* on the non-submodular edges. We formalize these intuitions in Sec. 3.3.

### 3.2. Defining the ORI Graph

When order reduction is applied to each factor in the HOMRF, new lower-order terms are generated. A given pairwise term (*e.g.*, the term  $x_1x_4$ ) could be associated with any number of factors in the original graph. Thus, the coefficients on pairwise terms are collected and summed over (possibly) many factors. In complicated graphs with many high-order factors, it is possible that a non-submodular pairwise term (say  $+5x_1x_4$ ) resulting from reducing a first factor (say  $x_1x_2x_4$ ) can be “cancelled out” when that same edge appears in submodular form with greater magnitude (say  $-8x_1x_4$ ) from the order reduction of a second factor (say  $x_1x_3x_4$ ).

As indicated in Section 2, each ORI-node  $v$  corresponds to a HOMRF factor  $A_v \in \mathcal{E}_{\mathcal{H}}$ , and takes a label  $o_v$  that decides which variable reduction option is selected from

among the reduction choices.

ORI-edges  $e \in \mathcal{E}_{\mathcal{G}}$ ,  $e = \{u, v\}$  represent the interaction between the order-reduction choices of nearby factors  $A_u, A_v \in \mathcal{E}_{\mathcal{H}}$  in the HOMRF. The ORI graph has an edge between ORI-nodes when the corresponding factors in the HOMRF share a pair of variables. More formally, let  $V_{ij} = \{A \in \mathcal{E}_{\mathcal{H}} \mid A \ni i, j\}$  be the set of all factors in the HOMRF that are associated with both variables  $x_i$  and  $x_j$  and therefore contribute to the  $x_ix_j$  term in the final PMRF as well. There is a ORI-edge between the members of  $V_{ij}$  whenever that set contains two or more members ( $|V_{ij}| \geq 2$ ).

Figure 3 shows an example HOMRF (a), along with the corresponding ORI graph (b), formed by replacing each factor with a node and creating edges for factors that share pairs of variables.

**Example.** Consider two factors in the HOMRF:  $A_u = \{1, 2, 3\}$  and  $A_v = \{2, 3, 5\}$ , where the first factor represents the interaction between  $x_1, x_2$  and  $x_3$  and the second factor captures the interaction between variables  $x_2, x_3$ , and  $x_5$ . Order reduction on the terms  $x_1x_2x_3$  and  $x_2x_3x_5$  introduce new variables  $x_6$  and  $x_7$ , respectively, and associated pairwise terms. In the final pairwise MRF  $\mathcal{P}$ , the coefficient in the polynomial expansion for the term  $x_2x_3$  depends not only on the parameters for that edge in the original HOMRF, but also on the reductions of both  $A_u$  and  $A_v$ , since both reductions emit an  $x_2x_3$  term.

In general, the energies for a term  $x_ix_j$  in a PMRF  $\mathcal{P}$  are found by summing over the emitted reduced factors from each of the HOMRF-factors containing  $i, j$ . Let  $\phi_{v, o_v}(i, j)$  be the coefficient associated with the term  $x_ix_j$  emitted from the  $v^{\text{th}}$  factor in the HOMRF, reduced according to the method  $o_v$ . The coefficient for a term  $x_ix_j$  in the final PMRF is:



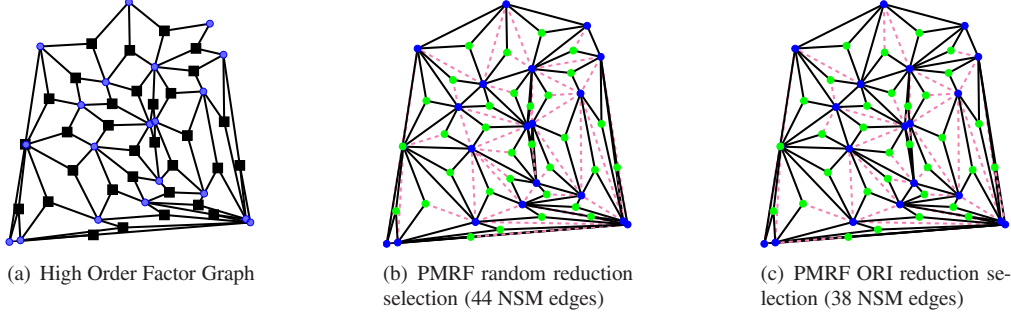


Figure 4. Order Reduction Inference can be used to produce a pairwise MRF with fewer non-submodular edges. (a) The original high-order MRF (HOMRF), with black squares to indicate factors, blue circles to indicate variables. In (b) and (c) solid black lines represent submodular edges, and pink dashed lines show non-submodular edges (NSMs). The PMRF resulting from ORI is shown in (c), where the global energy function is defined to minimize the number of NSMs. In this example, there are only 38 NSMs (c) versus 44 when random selection is used (b).

$$p_{ij}(\mathcal{O}) = \sum_{v \in V_{i,j}} \phi_{v,o_v}(i,j) \quad (11)$$

### 3.3. Selecting ORI Graph Parameters

We consider two different types of ORI-energy functions  $E_{\mathcal{G}}(\mathcal{O})$ : one for minimizing the number of non-submodular edges produced in the PMRF, and the other for minimizing the sum of the magnitude of the non-submodular terms in the PMRF.

**Minimizing non-submodular edge count:** The first energy function  $E_{1\mathcal{G}}(\mathcal{O})$  indicates the number of non-submodular edges in the PMRF resulting from a particular order reduction  $\mathcal{O}$ . Thus, by minimizing  $E_{1\mathcal{G}}(\mathcal{O})$ , we can choose an order reduction that produces as few non-submodular edges as possible. Ideally, if the PMRF could have no non-submodular edges, then exact inference is possible. However, this is unlikely to occur so we attempt to minimize the total number of non-submodular edges in the PMRF.

For simplicity of notation (and to match our current implementation), we will assume that a term in the PMRF  $x_i x_j$  is associated with at most two factors in the HOMRF (*i.e.* the set  $V_{i,j}$  contains at most two members). Let the members of  $V_{i,j}$  be  $A_u$  and  $A_v$ , representing the factors of the HOMRF associated with the PMRF term  $x_i x_j$ . Then  $E_{1\mathcal{G}}(\mathcal{O})$  is realized by defining the potentials on the ORI graph as follows:

$$\theta_{uv}(o_u, o_v) = \mathbb{I}_{[0 \infty)}(p_{ij}(o_u, o_v)) \quad \text{where,} \quad (12)$$

$$\mathbb{I}_S(x) = \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

Thus, the pairwise terms  $\theta_{uv}(o_u, o_v)$  are used to enumerate the number of non-submodular edges produced when factor reduction choices interact. Eq. (12) can be extended for defining high-order factors in the ORI graph when  $V_{i,j}$

contains more than two factors (*i.e.* when the term  $x_i x_j$  is associated with more than two factors in the HOMRF, in which case the ORI-graph itself would contain high-order terms.) In our current implementation of ORI, we force  $\mathcal{G}$  to be strictly pairwise by creating an ORI-edge for each pair of members of the set  $V_{i,j}$ . This limit in our implementation does not affect our experiments that use Delaunay triangulations over points in a plane, as it is geometrically impossible for an edge in the PMRF to be shared by more than two factors in the HOMRF.

The unary term for an ORI-node  $v$  counts the number of non-submodular edges that depend only on the specific order-reduction method  $o_v$  at the  $v^{\text{th}}$  factor of the HOMRF. To define this term, first we find  $U_v$ , the set of all pairwise terms in the PMRF, each of which is associated only with factor  $v$  of the HOMRF and not with any other factors of the HOMRF (*e.g.*, all pairwise terms containing an auxiliary variable corresponding to a single factor of the HOMRF). The members of  $U_v$  are all those edges  $(i, j)$  of the PMRF, such that  $V_{i,j}$  contains only the single element  $v$ . Then the unary term  $\theta_v(o_v)$  is defined as:

$$\theta_v(o_v) = \sum_{(i,j) \in U_v} \mathbb{I}_{[0 \infty)}(p_{ij}(o_v)). \quad (14)$$

Once the parameters are defined, TRW is employed to find the labeling on the ORI graph that approximately minimizes the energy function, and thereby produced a PMRF with fewer non-submodular edges.

Fig. 4 shows that this global energy function accomplishes the goal of selecting order reductions that produce a smaller number of non-submodular edges. A random selection of order reduction produces 44 non-submodular edges, while ORI's PMRF contains only 38 non-submodular edges.

In initial experiments, we applied ORI to 500 randomly generated HOMRFs over Delaunay graphs on 50 points (see Section 4 for details on construction). On average, ORI allowed us to produce PMRFs with fewer

non-submodular edges (8.6% reduction, with 117.1 non-submodular edges on average using ORI with  $E_{1G}(\mathcal{O})$  versus 128.1 non-submodular edges by randomly selecting a reduction). However, we found *no statistical difference* when approximate inference was applied to the resulting PMRFs. That is, approximate inference on the PMRF designed to have fewer non-submodular edges is no more accurate than inference on a PMRF produced by randomly selecting order-reduction methods.

**Minimizing the non-submodular magnitude:** For the second energy function  $E_{2G}(\mathcal{O})$ , we chose to minimize the total *magnitude* of the non-submodular terms. Fortunately, in contrast to the first energy function, the resulting PMRFs were found to facilitate approximate inference in this case.

In computing  $E_{2G}(\mathcal{O})$ , submodular edges in the PMRF incur no cost, but non-submodular edges incur a cost equal to the magnitude of the coefficient on the edge term. Note that this energy function does not explicitly limit the *number* of non-submodular edges, but rather attempts to keep their magnitude small.

Then  $E_{2G}(\mathcal{O})$  is realized by defining the potentials on the ORI graph as follows, again with  $u, v \in V_{ij}$ :

$$\theta_{uv}(o_u, o_v) = p_{ij}(o_u, o_v) \cdot \mathbb{I}_{[0, \infty)}(p_{ij}(o_u, o_v)). \quad (15)$$

The unary term  $\theta_v(o_v)$  is defined as:

$$\theta_v(o_v) = \sum_{i,j \in U_v} p_{ij}(o_v) \cdot \mathbb{I}_{[0, \infty)}(p_{ij}(o_v)) \quad (16)$$

As with  $E_{1G}(\mathcal{O})$ , we applied this method to 500 randomly generated HOMRFs over Delaunay graphs on 50 points. On average, ORI produced PMRFs with a smaller total weight on non-submodular edges by about 11.2% (a mean value of 122.4 using ORI with  $E_{2G}(\mathcal{O})$  versus 137.9 with a random reduction). Most importantly, we found that the performance of approximate inference algorithms *improved significantly* in PMRFs resulting from ORI, compared to random reductions.

## 4. Experiments

We performed experiments on synthetic data and real-world applications to evaluate the effectiveness of ORI.

### 4.1. HOMRFs on Synthetic Delaunay Graphs

We consider graph structures produced by Delaunay triangulation over  $N = \{50, 200, 1000\}$  random points in a 2D plane, where each point represents a variable in a high-order MRF. A high-order factor is formed over each triangle in the graph. In a manner similar to Kolmogorov [10], the parameters of the factors are randomly sampled from a Gaussian distribution. Although the problem parameters are synthetic, the topological structures are closely related

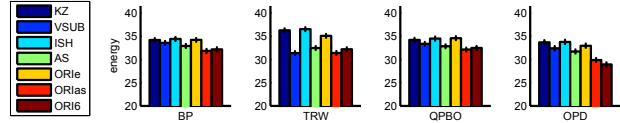


Figure 5. The mean energy for labellings on 500 HOMRFs from Delaunay triangulations over 50 points. The overall best performance is using ORI6 for order reduction and using OPD [1] for approximate inference. Among static order-reduction methods, the asymmetric (AS) reduction is competitive across all approximate inference methods.

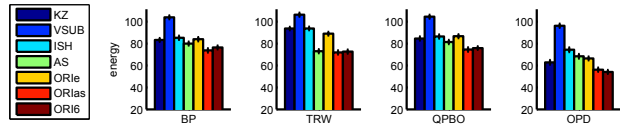


Figure 6. The mean energy for labellings on 500 HOMRFs from Delaunay triangulations over 200 points. Again, the overall best performance is using ORI6 for order reduction and using OPD [1] for approximate inference. Order reduction with ORI using either the asymmetric reductions (ORIs) or all six reductions (ORI6) generally have the best (smallest) energies.

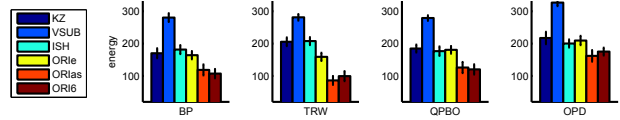


Figure 7. The mean energy for labellings on 10 HOMRFs from Delaunay triangulations over 1000 points. In this case, the overall best performance is using ORIs for order reduction and with TRW for approximate inference. For each approximate inference method, one of the ORI variants has the smallest energy.

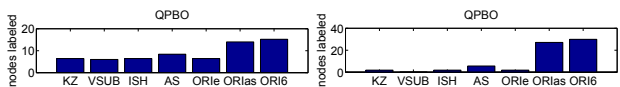


Figure 8. ORIs and ORI6 result in significant improvement in the average number of labeled nodes with QPBO for both D50 (left) and D200 (right). This is an indication that ORI produces an “easier” pairwise MRF than the other reduction methods.

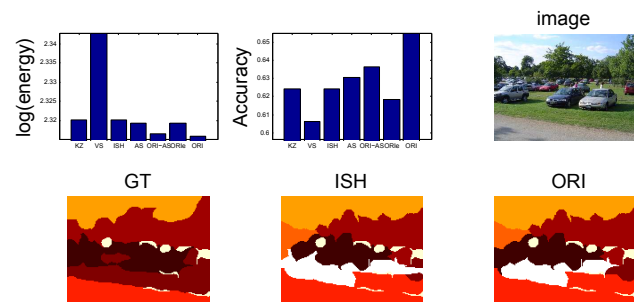


Figure 9. For an image from the MSRC set, we show ORI6 produces a labelling with the lowest energy (upper left), and the best accuracy (upper middle). The bottom row shows the resulting improved labelling from ORI (right) versus Ishikawa’s labelling (middle) versus the ground truth (left). The improvement is in the left part of the image.

to applications such as superpixel adjacency graphs and stereo matching over triangles [4].

We evaluate three variants of ORI: with ORle, each factor has the choice of being reduced with one of existing

techniques (KZ, VSUB or ISH); with ORIas, each factor has the choice of being reduced with one of the three variants of asymmetric reduction from §2.2; and with ORI6, each factor may be reduced by any of the six reduction choices. When ORIas is used, TRW finds the best order reduction according to the energy function  $\mathcal{E}_{2G}(\mathcal{O})$  over 98% of the time with the Delaunay ORI graphs.

For each order-reduction technique, the high-order MRF is reduced to a pairwise MRF that is then passed to one of a number of approximate inference engines. Note that exact inference cannot be performed efficiently whenever the resultant pair-wise MRF contains at least one non-submodular edge. We experiment with loopy belief propagation, tree-reweighted (TRW) message passing [24], QPBO [2,20] followed by TRW (for nodes that QPBO does not label), and outer-planar decomposition (OPD) [1].

Performing approximate inference on the pairwise MRF produces a label for each node, denoted as  $\mathcal{V}_P^*$ . The set of nodes of pairwise MRF is a superset of the variables in the original high-order MRF ( $\mathcal{V}_P \supseteq \mathcal{V}_H$ ), so  $\mathcal{V}_P^*$  provides an inferred label for each variable in the high-order MRF. Using  $\mathcal{V}_H^*$  from a particular order reduction and approximate inference algorithm, the model energy  $E(\mathcal{V}_H^*)$  of the original HOMRF can be found, which is comparable across reductions.

Each experiment is repeated over many graphs (500 for D50 and D200, and 10 for D1000), with each repetition having a new set of points from which Delaunay triangles are formed and new factor parameters occur. The results are shown in Figures 5, 6 and 7. We report our results using the mean energy of each labelling. We find that across all the inference engines, ORI achieves the lowest energies. We also note that for D50 and D200, OPD achieves significantly lower energies compared to the remaining inference engines.

Inference on the ORI graph requires extra processing time, but typically it is a smaller problem than the original HOMRF. In an exemplar D200 graph, the original HOMRF contains 200 nodes, 584 edges, and 385 triplet factors. The corresponding ORI graph has 385 nodes (one per triplet), 571 edges, and no triplet factors. Inference on the ORI graph required 2.6 ms (for 100 iterations of TRW-S), and inference on the resulting PMRF (585 nodes and 1739 edges) required 29.6 ms seconds; thus an overhead of 8.8%.

As other authors have noted [6], the number of nodes left unlabeled by QPBO is indicative of the “degree” of non-submodularity in the problem. Since the goal of ORI was to explicitly reduce this quantity, it is natural to verify if ORI results in QPBO labelling a larger portion of nodes in a graph. Figure 8 shows this to be true.

Reduction Method	Mean Energy	Mean Residual Energy	Best
KZ [12]	138.97	0.133	245
VSUB [17]	158.20	19.368	28
ISH [6]	138.94	0.108	247
ASYM §2.2	139.14	0.302	232
ORLe	138.92	0.082	248
ORIas	139.04	0.213	240
ORI6	<b>138.89</b>	<b>0.054</b>	<b>255</b>

Table 1. On 265 energy functions from MSRC images, ORI (with 6 reduction choices per factor) produces the best labelling of all reduction methods most often (255 times). In addition, ORI produces labellings with the smallest total energy, and the smallest mean residual energy (the mean between the energy for ORI’s labelling’s energy, and the energy of the best labelling found using any reduction method in this table). TRW was used for approximate inference.

## 4.2. Object Labelling

In our real-world application, we experiment with the task of image labelling, *i.e.*, labelling a superpixel from an image with one of 21 object classes, such as sky, grass, water, *etc.* We work with 265 test images from the MSRC dataset [22]. We model each test image with a high-order MRF where the variables in the graph correspond to superpixels in the image, and each variable has a 21-dimensional label space. On average, each image contains about 127 superpixels. The node potentials are discriminatively learned via TextonBoost [22] color and texture features.

Triplet cliques are formed by Delaunay triangulation over the centroids of the super-pixels of each image. The triplet clique energies are used to accomplish label smoothing and are the negative log of the triplet class co-occurrences matrix from a held-out training set. The relative weight of the node energies is manually set at 0.9 versus a weight of 0.1 for the smoothness terms, to avoid over-smoothing. Parameter learning in scenarios such as this is an active area of research in its own right, and could be used to establish the parameters for the HORMFs.

Since this label space is non-binary, we use alpha expansion. Within each alpha-expansion iteration, we reduce the higher-order MRF using ORI or other order-reduction techniques. The ORI inference is solved using TRW. The results are shown in Table 1. Overall, ORI6 has the best performance. It is worth noting that allowing ORI to select between multiple choices helps. For example, when ORI selects between KZ, VSUB, and ISH, the ORI result (ORLe) is at least as good as any of the individual constant reduction choices. The same is true for ORIas, which performs better than asymmetric reduction alone (and random selection of the reduction per factor). Figure 9 shows an image from the test set, the energy and labelling accuracies achieved by the different order-reduction techniques, as well as the set of labels determined by Ishikawa’s technique and ORI, compared to the ground-truth labelling.



Figure 10. An image from the geometric labeling dataset (left), the ground-truth labelling (center), and the inference result using ORIs for order reduction.

KZ	VSUB	ISH	ASYM	ORie	ORIs	ORI6
0.0682	6.71	0.0687	0.0446	0.0687	<b>0.0441</b>	0.0682

Table 2. On 300 images for the geometric labeling problem, ORIs has the lowest residual energy. Of the static reduction methods, ASYM from Sec. 2.2 performs best. TRW is used for approximate inference.

### 4.3. Geometric Surface Labelling

Finally, we also experimented with the geometric labelling problem. We follow the setup of Ramalingam *et al.* [16], who formulated the geometric labelling problem of Hoeim *et al.* [5] as an inference problem in HOMRFs. The goal is to label each superpixel as one of three classes — ground, vertical, or sky. We use the features and local classifiers of Hoeim *et al.* [5]. Node energies in this HOMRF are the confidence of the local classifiers. The graph-structure is an adjacency graph over superpixels. For edge energies we use negative log of co-occurrence counts from the training dataset. Following [16], we form triplets of nearly vertical columns of superpixels and use potentials that force these triplets to avoid bad labellings such as — ground, sky, ground — for three nearly vertical superpixels. Figure 10 shows the results of inference on one of the 300 test images. Table 2 shows the mean energy results from inference. Over the test set, ORIs has the lowest energy labellings. ORI6 performed about the same as KZ, and not as good as ORIs, possibly because of the larger label space of the ORI graph (6 vs. 3).

## 5. Discussion

In summary, this paper presents the idea that order reduction in HOMRFs should not be viewed in isolation of the inference procedure that will follow it. It should also not be viewed as a local problem that may be solved one high-order factor at a time. Instead, we present the idea that reductions can be selected to minimize a global property of the MRF. In essence, selecting good order-reduction methods for each factor in an MRF is a labelling problem in its own right. Our proposed Order Reduction Inference algorithm defines a global energy related to the magnitude of non-submodular terms in the reduced MRF, and then uses standard pairwise inference techniques to solve that labelling problem. In addition, we introduce a family of asymmetric reductions on second-order factors that have been shown to perform well within the ORI frame-

work. We show that ORI provides significant improvement for approximate inference across a wide range of approximate inference methods. Future work involves extending ORI to handle quartic and higher-order factors, which will increase the domain of vision problems for which it is applicable.

## References

- [1] D. Batra, A. 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 Appl. Math.*, 123(1-3):155–225, 2002.
- [3] D. Cremers and L. Grady. Statistical priors for efficient combinatorial optimization via graph cuts. In *ECCV*, 2006.
- [4] B. Glocker, H. Heibel, N. Navab, P. Kohli, and C. Rother. Triangle-flow: Optical flow with triangulation-based higher-order likelihoods. In *ECCV*, 2010.
- [5] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *IJCV*, 75(1), 2007.
- [6] H. Ishikawa. Higher-order clique reduction in binary graph cut. In *CVPR*, 2009.
- [7] H. Ishikawa. Transformation of general binary MRF minimization to the first order case. *PAMI*, 99(PrePrints), 2010.
- [8] J. Kappes and C. Schnrr. MAP-inference for highly-connected graphs with DC-programming. In *Pattern Recognition*, volume 5096 of *LNCS*, pages 1–10. Springer Verlag, 2008.
- [9] P. Kohli, L. Ladicky, and P. H. S. Torr. Robust higher order potentials for enforcing label consistency. *CVPR*, 2008.
- [10] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, 2006.
- [11] V. Kolmogorov and C. Rother. Comparison of energy minimization algorithms for highly connected graphs. In *ECCV*, pages II: 1–15, 2006.
- [12] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 2004.
- [13] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr. Graph cut based inference with co-occurrence statistics. In *ECCV*, 2010.
- [14] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *ICCV*, 2009.
- [15] S. Nowozin and C. Lampert. Global connectivity potentials for random field models. In *CVPR*, 2009.
- [16] S. Ramalingam, P. Kohli, K. Alahari, and P. H. S. Torr. Exact inference in multi-label CRFs with higher order cliques. *CVPR*, 2008.
- [17] I. Rosenberg. Reduction of bivalent maximization to the quadratic case. *Cahiers du Centre d'Etudes de Recherche Operationnelle*, 17:7174, 1975.
- [18] S. Roth and M. Black. Fields of experts. *IJCV*, 82(2), April 2009.
- [19] C. Rother, P. Kohli, W. Feng, and J. Jia. Minimizing sparse higher order energy functions of discrete variables. In *CVPR*, 2009.
- [20] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary MRFs via extended roof duality. In *CVPR*, June 2007.
- [21] D. Schlesinger and B. Flach. Transforming an arbitrary minsum problem into a binary one. *Technical Report TUD-FI06-01, Dresden University of Technology*, 2006.
- [22] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*, 2006.
- [23] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.
- [24] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second order smoothness priors. In *CVPR*, 2008.