

# Tensor power method for efficient MAP inference in higher-order MRFs

Dimitri Semenovitch, Arcot Sowmya  
 University of New South Wales  
 dimitris@cse.unsw.edu.au

## Abstract

*We present a new efficient algorithm for maximizing energy functions with higher order potentials suitable for MAP inference in discrete MRFs. Initially we relax integer constraints on the problem and obtain potential label assignments using higher-order (tensor) power method. Then we utilise an ascent procedure similar to the classic ICM algorithm to converge to a solution meeting the original integer constraints.*

## 1. Introduction

Many problems in computer vision and related fields such as segmentation, denoising, 3D reconstruction or graph matching can be reduced to performing inference in Conditional Random Field (CRF) or Markov Random Field (MRF) models. Conventionally pairwise potentials are employed as in this case many efficient algorithms for finding (approximate) Maximum a Posteriori (MAP) solutions are available such as Belief Propagation (BP) [4], graph cuts [2] and tree reweighted message passing [8]. Such models, however, have limited flexibility and expressive power as they can only model interactions between pairs variables.

Employing higher order potentials on the other hand may allow to capture multiple interactions and better model natural statistics of images. While it has been shown that many vision tasks could benefit from the use of such models [10, 15] they have not gained wide acceptance because of difficulties with inference, both obtaining good approximations and prohibitive computational costs.

A limited number of algorithms for inference in higher order MRFs has been described to date. Among these are extensions to BP to enable more efficient optimization of a particular class of energy functions with higher order cliques [14]. Kohli et al. [7] extended the graph cut approach to a different class of higher order energies ( $P^n$  Potts models). Other approaches seek

to address the problem by transforming higher order MRFs to models with pairwise terms only and applying standard techniques [5, 9].

In this work we present a simple new algorithm for maximizing energy functions with higher order potentials suitable for MAP inference in discrete MRFs. In contrast with most of the recent work our algorithm is suitable for arbitrary graphs and does not rely on restricting the class of admissible energy functions. Tasks such as scene annotation can benefit from employing graphs with arbitrary higher order potentials, for example, in the CRF framework. In our approach we initially relax integer constraints on the problem, following [12] and obtain potential label assignments using higher-order (tensor) power method. Then we utilise an ascent procedure similar to the classic iterated conditional modes (ICM) [1] algorithm to converge to a solution meeting the original integer constraints.

## 2 Problem formulation

We can represent an energy function with discrete states and third order potentials in the following form (for probability densities this would require a log transform), due to somewhat clumsy notation we do not consider the general case until later:

$$E = \sum_{ia;jb;kc} H_{ia;jb;kc} x_{ia} x_{jb} x_{kc} \quad (1)$$

Here  $H_{ia;jb;kc}$  represents the third order term describing how well the label  $a$  at site  $i$  agrees with the label  $b$  at site  $j$  and label  $c$  at site  $k$ . It can depend on other data in the case of a CRF model or, for example, represent a data independent smoothness constraint. If lower order terms are required, they can be rolled up into  $H$ . We set  $H_{ia;jb;kc} = 0$  if the three sites do not form a clique or if different labels are assigned to the same site (e.g. if  $i = j$  and  $a \neq b$ ).

Now  $x$  is an indicator vector with an entry for each permitted combination of sites and labels, such that

$x_{ia} = 1$  if site  $i$  is assigned label  $a$  and  $x_{ia} = 0$  otherwise. Note that  $x$  has dimension  $nm$  where  $n$  is the number of sites and  $m$  the number of labels for each site (we assume that all sites have the same number of labels without the loss of generality). There is a further constraint that each site can be given only one label. We can then express the energy maximisation problem as finding the vector  $\hat{x}_{opt}$  such that:

$$\hat{x}_{opt} = \operatorname{argmax}_x \sum_{ia;jb;kc} H_{ia;jb;kc} x_{ia} x_{jb} x_{kc} \quad (2)$$

subject to the constraints on  $x$  outlined above:

$$x \in \{0, 1\}^{nm}, \quad \sum_a x_{ia} = 1 \quad (3)$$

## 2.1 Tensor formalism

The above problem can be naturally formulated in the language of tensors. While more conceptual definitions are available here we take a tensor of order  $N$  to be an  $M_1 \times M_2 \times \dots \times M_N$  dimensional  $N$ -way array. For example, a scalar is a tensor of order zero, a vector is a tensor of order one and a matrix is a tensor of order two. An order  $N$  supersymmetric tensor is a an  $N$ -way array whose elements are invariant under any permutation of their indices. We can see that  $\mathcal{H}$  in (1) is a supersymmetric tensor of order three with entries  $H_{ia;jb;kc}$

Following [6] we define the  $n$ -mode matrix unfolding,  $\mathcal{T}_{(n)}$ , of an  $M_1 \times M_2 \times \dots \times M_N$  tensor  $\mathcal{T}$  of order  $N$  with entries  $\mathcal{T}_{i_1;i_2;\dots;i_N}$  is defined as the  $M_n \times M_1 M_2 \dots M_{n-1} M_{n+1} \dots M_N$  matrix whose columns are the  $M_n$ -dimensional vectors obtained from  $\mathcal{T}$  by varying the index in dimension  $n$  and keeping the other indices fixed. It can be seen that for a supersymmetric tensor  $\mathcal{T}$  all  $n$ -mode matrix unfoldings are equal,  $\mathcal{T}_{(1)} = \mathcal{T}_{(2)} = \dots = \mathcal{T}_{(N)}$ .

We can now rewrite the equation (1) as:

$$E = x^T \mathcal{H}_{(1)}(x \otimes x) \quad (4)$$

where  $\otimes$  denotes the Kronecker product, which for two  $n$  dimensional vectors  $a$  and  $b$  is defined as an  $n^2$  dimensional vector:

$$a \otimes b = \begin{pmatrix} a_1 b \\ a_2 b \\ \vdots \\ a_n b \end{pmatrix} \quad (5)$$

In the new notation we can formulate the general optimization problem as:

$$\hat{x}_{opt} = \operatorname{argmax}_x x^T \mathcal{H}_{(1)}(x \otimes x \otimes \dots \otimes x) \quad (6)$$

subject to constraints in (3).

## 2.2 Higher order power method

To make finding a solution to the integer optimisation problem (8) feasible we relax the constraints imposed on  $x$  in (3), replacing them with:

$$x \in \mathbb{R}^{nm}, \quad \|x\|_2 = 1 \quad (7)$$

We can obtain  $x$  satisfying the relaxed constraints as follows:

$$\hat{x}_{rel} = \operatorname{argmax}_x \frac{x^T \mathcal{H}_{(1)}(x \otimes x \otimes \dots \otimes x)}{\|x\|} \quad (8)$$

In the case of the second order energy function the above expression becomes the well known Rayleigh quotient which is maximized by the leading eigenvector of  $H$ .

---

### Algorithm 1 Power method

---

**Require:** Square matrix  $H$

- 1: Initialize  $x$  at random.
- 2: **repeat**
- 3:    $x \leftarrow Hx$
- 4:    $x \leftarrow \frac{x}{\|x\|}$
- 5: **until** convergence

**Output:** Vector  $x$

---

The power method is a simple algorithm for computing the leading eigenvector of a matrix. This algorithm efficiently converges to the leading eigenvector of the input matrix provided there is a sufficient gap between its largest and second largest eigenvalues. If  $H$  is sparse each iteration of the power method algorithm requires only  $O(k)$  operations, where  $k$  is the number of non-zero elements in  $H$ .

To find the maximum of the general higher order energy function in (8) we apply a generalization of the power method suggested in [11], initially for computing low rank approximations of a supersymmetric tensor  $\mathcal{H}$ . This algorithm was also recently used in [3].

While this algorithm is not guaranteed not reach a global maximum, it converges to a local maximum [6] for tensors which result in convex functions of  $x$ . Experimentally it converges almost always to a good solution.

---

**Algorithm 2** Higher order power method (HOPM)

---

**Require:** Supersymmetric tensor  $\mathcal{H}$ .

- 1: Initialize  $x$  at random.
- 2: **repeat**
- 3:  $x \leftarrow \mathcal{H}_{(1)}(x \otimes x \otimes \dots \otimes x)$
- 4:  $x \leftarrow \frac{x}{\|x\|}$
- 5: **until** convergence

**Output:** Vector  $x$ 

---

If the entries of  $\mathcal{H}$  are all positive the resulting solution vector will have only positive values. This will allow us to interpret the entries of  $x$  as the degree of confidence of a particular label assignment.

### 2.3 Enforcing integer constraints

Once we have obtained a solution to the optimization problem (8) with relaxed constraints, we still need to discretize the vector  $x$  respecting the original constraints (3). To do so while also improving the quality of the label assignments we propose a modification of the higher order power method which is also closely related the classical ICM algorithm [1] for MRF inference (see [13] for another recent extension of ICM).

---

**Algorithm 3** Discretising the initial solution (ICM)

---

**Require:** Supersymmetric tensor  $\mathcal{H}$ .

- 1: Initialize  $x$  using HOPM (algorithm 2).
- 2:  $y \leftarrow \pi(x)$
- 3:  $S^* \leftarrow y^T \mathcal{H}_{(1)}(y \otimes y \otimes \dots \otimes y)$
- 4: **repeat**
- 5:  $x \leftarrow \pi(\mathcal{H}_{(1)}(y \otimes y \otimes \dots \otimes y))$
- 6:  $S \leftarrow x^T \mathcal{H}_{(1)}(x \otimes x \otimes \dots \otimes x)$
- 7: **if**  $S > S^*$  **then**
- 8:  $S^* \leftarrow S$ ,  $y \leftarrow x$
- 9: **end if**
- 10: **until**  $S \leq S^*$

**Output:** Vector  $y$ 

---

It essentially replaces the step 4 in the algorithm 2,  $x \leftarrow \frac{x}{\|x\|}$ , which is a projection to the unit sphere, with a projection to the solution set defined by (3). Such a projection can be defined as follows:

$$\pi(x) = \operatorname{argmax}_y y^T \mathcal{H}_{(1)}(x \otimes \dots \otimes x) \quad (9)$$

where  $y$  is a discrete vector satisfying (3). In our case it amounts to setting  $y_{ia}$  as follows for each site  $i$ :

$$y_{ia} = \begin{cases} 1, & a = \operatorname{argmax}_b x_{ib} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

The proposed algorithm is a modification of ICM where the updates are performed in parallel rather than sequentially. It can be thought of as a series of independent labeling problems where the next solution is found based on the current one. In our experiments the algorithm converged after several iterations, offering a considerable improvement over the straight forward discretisation of the output of the higher order power method.

### 2.4 Mixing different order potentials

When dealing with energy functions of mixed order, one solution is to include all the information into the tensor of the highest order potential. It can be easier, however, to employ the following simple extension of the algorithm 2:

---

**Algorithm 4** Higher order (mixed) power method

---

**Require:** Tensors  $\mathcal{H}$  and  $\mathcal{J}$ , matrix  $K$ .

- 1: Initialize  $x$  at random.
- 2: **repeat**
- 3:  $x \leftarrow \mathcal{H}_{(1)}(x \otimes x \otimes x \otimes x) + \mathcal{J}_{(1)}(x \otimes x) + Kx$
- 4:  $x \leftarrow \frac{x}{\|x\|}$
- 5: **until** convergence

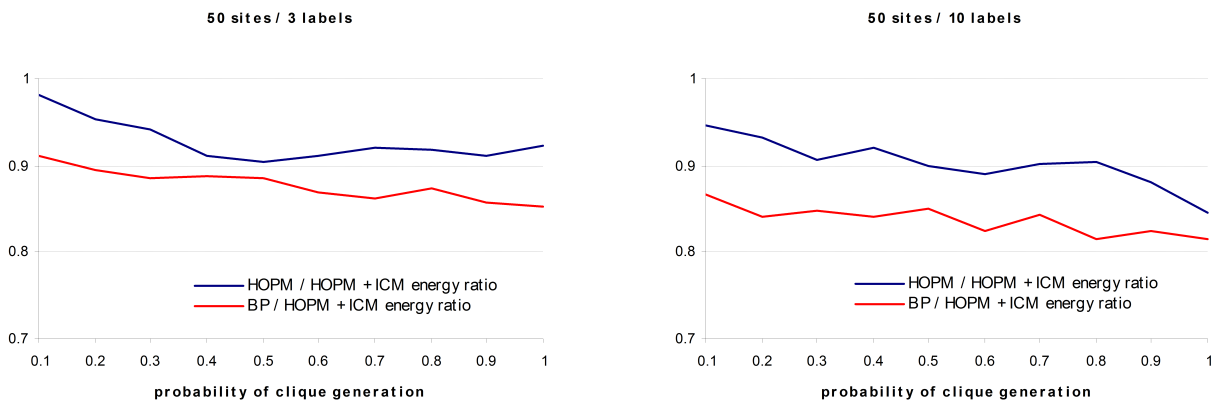
**Output:** Vector  $x$ 

---

It is also straight forward to modify algorithm 3 to use in this setting.

## 3. Experiments

We evaluated our algorithm against loopy belief propagation (max product) which is a standard algorithm for MAP estimation. Evaluation was performed using randomly generated third order energy functions, controlling the degree of connectedness and the number of labels. We generated the structure by assigning a group of sites to the same clique with a certain probability  $p_{clique}$ . Next we randomly grouped the sites into a number of disjoint sets. For cliques of connected sites  $(i, j, k)$ , and the uniform random variable  $p$ , we set  $H_{ia;jb;kc} = -\log(p)$ ,  $p \in [\epsilon, 1]$  with probability  $p_0$  and  $H_{ia;jb;kc} = 0$  otherwise. If  $a = b = c = 1$  and the sites  $(i, j, k)$  are in the same set we have  $p_0 = 0.4$ , otherwise  $p_0 = 0.2$ . This forces the third order potentials between



**Figure 1.** Average scores of HOPM and BP normalized by the score of HOPM+ICM over 50 experiments for different probabilities of click generation. On the left results for third order 50 site energy function with 3 labels and on the right for a third order 50 site energy function with 10 labels.

”correct” labels for connected sites in the same set to be on larger on average than the other potentials.

In Figure 1. we report the energies for the HOPM and BP normalized by the energies for HOPM+ICM for graphs of varying density. We see that HOPM+ICM reliably outperforms BP and that the ICM discretization step further improves performance. HOPM experiments used naive discretisation equivalent to a single iteration of ICM.

## 4. Conclusion

We have presented an efficient algorithm which approximates arbitrary higher order energy functions. Our experiments show that the proposed approach reliably outperforms standard loopy belief propagation and demonstrate the effectiveness of the proposed discretisation procedure. Future work may include modifying the basic algorithm to deal with continuous or ordered labels.

## References

- [1] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, 48:259–302, 1986.
- [2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 2001.
- [3] O. Duchenne, F. Bach, I. Kweon, and J. Ponce. A tensor-based algorithm for high-order graph matching. In *CVPR*, 2009.
- [4] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70(1), 2006.
- [5] H. Ishikawa. Higher-order clique reduction in binary graph cut. In *CVPR*, 2009.
- [6] E. Kofidis, Phillip, and A. Regalia. On the best rank-1 approximation of higher-order supersymmetric tensors. *SIAM Journal of Matrix Analysis and Applications*, 23:863–884, 2002.
- [7] P. Kohli, L. Ladicky, and P. H. S. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008.
- [8] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 2006.
- [9] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 2004.
- [10] X. Lan, S. Roth, D. Huttenlocher, and M. Black. Efficient belief propagation with learned higher-order markov random fields. In *ECCV*, 2006.
- [11] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. On the best rank-1 and rank-( $r_1, r_2, \dots, r_n$ ) approximation of higherorder tensors. *SIAM Journal of Matrix Analysis and Applications*, 21:13241342, 2000.
- [12] M. Leordeanu and M. Hebert. Efficient map approximation for dense energy functions. In *ICML*, 2006.
- [13] M. Leordeanu, M. Hebert, and R. Sukthankar. An integer projected fixed point method for graph matching and map inference. In *NIPS*, 2009.
- [14] B. Potetz. Efficient belief propagation for vision using linear constraint nodes. In *CVPR*, 2007.
- [15] O. Woodford, P. Torr, I. Reid, and A. F. . Global stereo reconstruction under second order smoothness priors. In *CVPR*, 2008.