This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the version available on IEEE Xplore.

Truncated Max-of-Convex Models

Pankaj Pansari
University of Oxford
The Alan Turing Institute
pankaj@robots.ox.ac.uk

M. Pawan Kumar
University of Oxford
The Alan Turing Institute
pawan@robots.ox.ac.uk

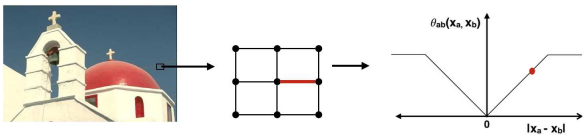
Abstract

Truncated convex models (TCM) are a special case of pairwise random fields that have been widely used in computer vision. However, by restricting the order of the potentials to be at most two, they fail to capture useful image statistics. We propose a natural generalization of TCM to high-order random fields, which we call truncated max-of-convex models *(TMCM). The energy function of TMCM consists of two types of potentials: (i) unary potential, which has no restriction on its form; and (ii) clique potential, which is the sum of the m largest truncated convex distances over all label pairs in a clique. The use of a convex distance function encourages smoothness, while truncation permits discontinuities in the labeling. By using $m > 1$, TMCM provides robustness towards errors in the definition of the cliques. To minimize the energy function of a TMCM over all possible labelings, we design an efficient st-MINCUT based range expansion algorithm. We prove the accuracy of our algorithm by establishing strong multiplicative bounds for several special cases of interest. Using standard real data sets, we demonstrate the benefit of our high-order TMCM over pairwise TCM, as well as the benefit of our range expansion algorithm over other st-MINCUT based approaches.*

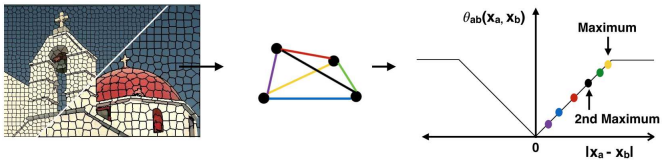
1. Introduction

Truncated convex models (TCM) are a special case of pairwise random fields that have been widely used for low-level vision applications. A TCM is defined over a set of random variables, each of which can be assigned a value from a finite, discrete and ordered label set. In addition, a TCM also specifies a neighborhood relationship over the random variables. An assignment of values to all the variables is referred to as a labeling. In order to quantitatively distinguish the labelings, a TCM specifies an energy function that consists of unary and pairwise potentials.

Given an input, the output is obtained by minimizing the energy function of a TCM over all possible labelings. While this is an NP-hard problem, several approximate algorithms have been proposed in the literature [3, 4, 12, 17, 20, 22, 24, 27, 30], which provide accurate solutions in practice [28].



(a) Truncated convex model



(b) Truncated max-of-convex model

Figure 1: *TMCM as generalization of TCM. In (a), given an image, TCM considers pairwise 4-neighborhood relationships and uses truncated convex distance function for pairwise potential. In (b), TMCM considers superpixels as cliques. The clique potential for m =2 is the sum of the first and second maximum over all the pairwise truncated convex distances such that no variable is used more than once.*

Since we cannot reasonably expect to improve the optimization of TCM, any failure cases must be addressed by modifying the model itself to better capture image statistics. To this end, we propose to address one of the main deficiencies of TCM, namely, the restriction to potentials of order at most two. Specifically, we propose a natural generalization of TCM to high-order random fields, which we refer to as *truncated max-of-convex models* (TMCM). Similar to TCM, our model places no restrictions on the unary potentials. Furthermore, unlike TCM, it allows us to define clique potentials over an arbitrary number of random variables. The value of the clique potential is proportional to the sum of the m largest truncated convex distances computed over disjoint pairs of random variables in the clique. Here, disjoint pairs imply that the label of no random variable is used more than once to compute the value of the clique potential. Figure 1 demonstrates how TMCM differs from TCM. The exact mathematical form of the TMCM energy function will be presented in section 4. The term m is a positive integer that is less than or equal to half the number of variables in the smallest clique. Importantly, the constant

of proportionality for each clique potential can depend on the input corresponding to all the random variables in the clique. This can help capture more interesting image statistics, which in turn can lead to a more desirable output. For example, in image denoising, the clique weights can depend on the variance of intensity over a superpixel (group of pixels with similar semantic and perceptual characteristics).

In order to enable the use of TMCMs in practice, we require an efficient and accurate energy minimization algorithm that can compute the output for a given input. To this end, we extend the range expansion algorithm for TCM to deal with arbitrary sized clique potentials. Our algorithm retains the desirable property of iteratively solving an st-MINCUT problem over an appropriate directed graph (where the number of vertices and arcs grows linearly with the number of random variables and cliques, and quadratically with the number of labels). As the st-MINCUT problem lends itself to several fast algorithms [2], this makes our overall approach computationally efficient. Furthermore, we provide strong theoretical guarantees on the quality of the solution for several special cases of interest, which establishes its accuracy. Our multiplicative bounds are better than the baselines for cases where comparison is possible. Using standard real data sets, we show the benefit of high-order TMCM over pairwise TCM, as well as the advantage of our range expansion algorithm over other st-MINCUT based approaches. The supplementary material as well as the latest version of the paper is available at https://arxiv.org/abs/1512.07815.

2. Related Work

Pairwise TCM offer a natural framework to capture low-level cues for vision problems such as image denoising, stereo correspondence, segmentation and optical flow [28]. However, the restriction to pairwise potentials limits their representational power.

For the past few years, there has been a growing interest in high-order models. Though other inference algorithms such as message passing are possible [9, 18, 21, 26, 29, 31], in this work our focus is on models that admit efficient st-MINCUT based solutions and provide strong theoretical guarantees on the quality of the solution. One early work was the P^n Potts model [15], which encourages label consistency over a set of random variables. This work was extended in [16], which introduced robustness in the P^n Potts model by taking into account the number of random variables that disagreed with the majority label of a clique. Both the P^n Potts model and its robust version lend themselves to efficient optimization via the expansion algorithm [3], which solves one st-MINCUT problem at each iteration. The expansion algorithm provides multiplicative bounds [11], which measure the quality of the estimated labeling with respect to the optimal one. Our work generalizes both the models, as well as the corresponding expansion algorithm. Specifically, when the truncation factor of our models is set to 1, we recover the robust P^n model. Furthermore, a suitable setting of the range expansion algorithm (setting the interval length to 1) recovers the expansion algorithm.

Jegelka and Bilmes [14] introduced a nonsubmodular high-order model which is based on edge cooperation and is optimizable using st-MINCUT, but the algorithm has weak approximation bounds. Delong et al. [6, 7] proposed a clique potential based on label costs that can also be handled via the expansion algorithm. However, unlike the robust P^n Potts model, their model provides additive bounds that are not invariant to reparameterizations of the energy function. This theoretical limitation is addressed by the recent work of Dokania and Kumar [8] on parsimonious labeling. Here, the clique potentials are defined as being proportional to a diversity function of the set of unique labels present in the clique labeling. Our work can be thought of as being complementary to parsimonious labeling. Specifically, while parsimonious labeling is an extension of pairwise metric labeling to high-order models, our work is an extension of TCM. The only metric that also results in a TCM is the truncated linear distance. As our experiments will demonstrate, our specialized range expansion algorithm provides significantly better results for truncated max-of-linear models compared to the hierarchical st-MINCUT approach of [8].

We note that there have been several works that deal with more general high-order potentials and design st-MINCUT style solutions for them. For example, Fix et al. [10] use the submodular max-flow algorithm [19], while Arora et al. [1] use generic cuts. However, the resulting algorithms are exponential in the size of the cliques, which prevents their use in applications that require very high-order cliques (with hundreds or even thousands of random variables). A notable exception to this is the work of Ladicky et al. [25], who proposed a co-occurrence based clique potential whose only requirement is that it should increase monotonically with the set of unique labels present in the clique labeling. However, the use of such a general clique potential still results in an inaccurate energy minimization algorithm, as will be seen in our experimental comparison.

3. Truncated Convex Models

A TCM is a random field defined by a set of discrete random variables $\mathbf{X} = \{X_a, a \in \mathcal{V}\}$, and a neighborhood relationship $\mathcal{E}$ over them (that is, X_a and X_b are neighboring random variables if $(a, b) \in \mathcal{E}$). Each random variable can take a value from a finite label set $\mathbf{L}$, which is assumed to be ordered to enable the use of convex distance functions. Without loss of generality, we define $\mathcal{V} = \{1, 2, \cdots, n\}$ and $\mathbf{L} = \{1, 2, \cdots, h\}$.

An assignment of values to all the random variables $\mathbf{x} \in \mathbf{L}^n$ is referred to as a labeling. To quantitatively distinguish the h^n possible labelings, a TCM defines an energy function that consists of two types of potentials. First,

the unary potential $\theta_a(x_a)$ that depends on the label x_a of one random variable X_a. Second, the pairwise potential $\theta_{ab}(x_a, x_b)$ that depends on the labels x_a and x_b of a pair of neighboring random variables (X_a, X_b). There are no restrictions on the form of the unary potentials. However, the pairwise potentials are defined using a truncated convex distance function over the label set.

To provide a formal specification of the pairwise potentials, we require some definitions. We denote a convex distance function by $d : \mathbb{Z} \rightarrow \mathbb{R}$ (where $\mathbb{Z}$ is the set of integers and $\mathbb{R}$ is the set of real numbers). Recall that a convex distance function satisfies the following properties: (i) $d(y) \geq 0$ for all $y \in \mathbb{Z}$ and $d(0) = 0$; (ii) $d(y) = d(-y)$ for all $y \in \mathbb{Z}$; and (iii) $d(y+1) - 2d(y) + d(y-1) \geq 0$ for all $y \in \mathbb{Z}$. Note that the above properties also imply that $d(y) \geq d(z)$ if $|y| \geq |z|$, for all $y, z \in \mathbb{Z}$. Examples of convex distance functions include the linear (that is, $d(y) = |y|$) and the quadratic distance function (that is, $d(y) = y^2$) distance.

Given two labels $l_i, l_j \in \mathbf{L}$, we use a convex function $d(\cdot)$ to measure the distance between them as $d(l_i - l_j)$, thereby encouraging smooth labelings. In order to prevent the overpenalization of the discontinuities in an image, it is common practice to truncate the convex distance function [3, 23, 30]. Formally, a truncated convex function is defined as $\min\{d(\cdot), M\}$, where M is the truncation factor. We now define the pairwise potential as $\theta_{ab}(x_a, x_b) = \omega_{ab} \min\{d(x_a - x_b), M\}$, where ω_{ab} is a (data-dependent) non-negative constant of proportionality.

Hence, a TCM specifies an energy function $E(\cdot)$ over the labelings $\mathbf{x} \in \mathbf{L}^n$ as follows:

$$E(\mathbf{x}) = \sum_{a \in \mathcal{V}} \theta_a(x_a) + \sum_{(a,b) \in \mathcal{E}} \omega_{ab} \min\{d(x_a - x_b), M\}. \quad (1)$$

Given an input (which provides the values of the unary potentials and the edge weights), the desired output is obtained by solving the following optimization problem: $\min_{\mathbf{x} \in \mathbf{L}^n} E(\mathbf{x})$. While this optimization problem is NP-hard, we can obtain an accurate approximate solution by using the efficient range expansion algorithm [23], as well as several other approaches based on st-MINCUT [3, 12, 22, 30] and linear programming [4, 17, 20].

4. Truncated Max-of-Convex Models

We now present a natural generalization of TCM to high-order random fields, which define potentials over random variables that form a clique (where all the random variables in a clique are neighbors of each other). Importantly, we do not place any restriction on the size of the clique.

Truncated Max-of-Convex Potentials. Consider a high-order clique consisting of the random variables $\mathbf{X_c} = \{X_a, a \in \mathbf{c} \subseteq \mathcal{V}\}$. We denote a labeling of the clique as $\mathbf{x_c} \in \mathbf{L}^c$, where we have used the shorthand $c = |\mathbf{c}|$ to denote the size of the clique. In order to specify the value of the clique potential for the labeling $\mathbf{x_c}$ we require a sorted list of the (not necessarily unique) labels present in $\mathbf{x_c}$. We denote this sorted list by $\mathbf{p}(\mathbf{x_c})$ and access its i-th element as $p_i(\mathbf{x_c})$. For example, consider a clique consisting of random variables $\mathbf{X_c} = \{X_1, X_2, X_3, X_4, X_5, X_6\}$. If the number of labels is $h = 10$, then one of the putative labelings of the clique is $\mathbf{x_c} = \{3, 2, 1, 4, 1, 3\}$ (that is, X_1 takes the value 3, X_2 takes the value 2 and so on). For this labeling, $\mathbf{p}(\mathbf{x_c}) = \{1, 1, 2, 3, 3, 4\}$. The value of $p_1(\mathbf{x_c})$ and $p_2(\mathbf{x_c})$ is 1, the value of $p_3(\mathbf{x_c})$ is 2 and so on. Given a convex function $d(\cdot)$, a truncation factor M and an integer $m \in [0, \lfloor c/2 \rfloor]$, the clique potential $\theta_{\mathbf{c}}(\cdot)$ is defined as

$$\theta_{\mathbf{c}}(\mathbf{x_c}) = \omega_{\mathbf{c}} \sum_{i=1}^{m} \min\{d(p_i(\mathbf{x_c}) - p_{c-i+1}(\mathbf{x_c})), M\}. \quad (2)$$

Here, $\omega_{\mathbf{c}} \geq 0$ is the clique weight that does not depend on the labeling. However, it can be chosen based on the observed data - for instance, we may want to assign small weights to cliques with large variance of intensity or disparity. The term inside the summation is the truncated value of the i-th largest distance between the labels of all pairs of random variables within the clique, subject to the constraint that the label of no random variable is used more than once in the computation of the clique potential value. In other words, our clique potential is proportional to the sum of the truncation of the m largest convex distance functions over disjoint pairs of random variables.

As an example, consider a clique of size 6 assigned the labeling {1, 2, 3, 4, 5, 6} and M = 3, m = 3, $\omega_{\mathbf{c}}$ = 1. According to equation (2), $\theta_{\mathbf{c}}(\mathbf{x_c})$ = min(6 - 1, 3) + min(5 - 2, 3) + min(4 - 3, 3) = 3 + 3 + 1 = 7. Using this and other instances, table 1 demonstrates how TMCM ensures smooth labelings, prevents overpenalization of discontinuities (desirable at object boundaries) and provides robustness to erroneous clique definitions that may happen if division of image into superpixels is not perfect.

Given an input, the desired output is obtained by solving the following optimization problem: $\min_{\mathbf{x} \in \mathbf{L}^n} E(\mathbf{x})$. Note that TMCM is a generalization of the P^n Potts model [15] (m = 1, M = 1) as well as its robust version [16] ($m > 1$, M = 1). Furthermore, it is complementary to the recently proposed parsimonious labeling, which generalizes metric labeling. Henceforth, we assume the unary potentials are non-negative. This assumption is not restrictive as we can always add a constant to the unary potentials of a random variable. This modification would only result in the energy of all labelings changing by the same constant. As we shall see, our algorithm as well as its theoretical guarantees are invariant to such changes in the energy function.

Labeling	$m = 1$	$m = 2$	$m = 3$
(a) {1,1,1,1,2,2}	1	2	2
{1,2,3,4,5,6}	3	6	7
(b) {1,1,1,9,9,9}	3	6	9
{1,1,1,8,8,9}	3	6	9
(c) {1,1,1,1,1,7}	3	3	3
{1,1,1,2,3,4}	3	5	6

Table 1: *Clique potential value $\theta_{\mathbf{c}}(\mathbf{x_c})$ defined by a linear function with M = 3 and $\omega_{\mathbf{c}}$ = 1 for various values of m. Since clique size is 6, $0 \leq m \leq 3$. In pair (a), the first labeling is smoother than the second. Use of the largest convex distances assigns lower penalty to the smooth labeling. In the labelings of (b), one group of variables has low label and another high. In both cases, truncation ensures that the discontinuity is not overpenalized. The first labeling of (c) has a random variable with a very high label, possibly due to incorrect clique definition. When $m > 1$ is used, the presence of the erroneous variable is not heavily penalized.*

5. Optimization via Range Expansion

As TMCM is a generalization of TCM, it follows that the corresponding energy minimization problem is NP-hard. However, we show that the efficient and accurate range expansion algorithm can be extended to handle this more general class of energy functions.

Algorithm 1 shows the main steps of range expansion. The algorithm starts by assigning the random variables to an initial label (step 1). For example, all the random variables could be assigned to the label 1. Next, it selects an interval of consecutive labels of size at most h' (steps 3-4), where h' is specified as an input to the algorithm. We will see later in the section that the value of h' can be chosen to obtain the optimal worst case bound for specific instances of the TMCM. Next, it minimizes the energy over all the labelings that either allow a random variable to retain its current label, or choose a new label in the selected interval (step 5). If the energy of the new labeling is lower than that of the current labeling, then the solution is updated (steps 6-8). This process is repeated for all the intervals of consecutive labels of size at most h'. The entire algorithm stops when the energy cannot be reduced further for any choice of the interval.

The crux of the range expansion algorithm is problem (3), which needs to be solved for any given interval $\mathbf{I}$ and current labeling $\hat{\mathbf{x}}$. Unfortunately, this problem itself is NP-hard for TMCM. Indeed, when $h' = h$, problem (3) is equivalent to the original energy minimization problem. In order to operationalize the range expansion algorithm, we need to devise an approximate algorithm for problem (3). We achieve this in two steps. First, we obtain an overestimate of the energy function $E(\cdot)$, which we denote by $E'(\cdot)$. The energy function $E'(\cdot)$ is restricted to the labels in the interval $\mathbf{I}$ together with the labels specified by the current labeling $\hat{\mathbf{x}}$. Second, we minimize the overestimated energy $E'(\cdot)$ over all of its putative labelings by solving an equivalent st-MINCUT problem. We describe our two-step algorithm in the next two subsections in detail. Specifically, subsection 5.1 describes the exact form of the energy function $E'(\cdot)$, while subsection 5.2 describes the construction of the directed graph over which we solve the st-MINCUT problem to obtain the labeling $\mathbf{x}'$.

Algorithm 1 The range expansion algorithm.

input Energy function $E(\cdot)$, initial labeling $\mathbf{x}^0$, interval length h'.

1: Initialize the output labeling $\hat{\mathbf{x}} = \mathbf{x}^0$.
2: **repeat**
3: **for all** $i_m \in [-h' + 2, h]$ **do**
4: Define an interval of labels $\mathbf{I} = \{f, \cdots, l\}$ where $f = \max\{i_m, 1\}$ and $l = \min\{i_m + h' - 1, h\}$.
5: Obtain a new labeling $\mathbf{x}'$ by solving the following optimization problem:

$$\mathbf{x}' = \underset{\mathbf{x}}{\operatorname{argmin}}\, E(\mathbf{x}), \quad \text{s.t. } x_a \in \mathbf{I} \cup \{\hat{x}_a\}, \forall a \in \mathcal{V}. \tag{3}$$

6: **if** $E(\hat{\mathbf{x}}) > E(\mathbf{x}')$ **then**
7: Update $\hat{\mathbf{x}} = \mathbf{x}'$.
8: **end if**
9: **end for**
10: **until** The labeling does not change for any value of i_m.

output The labeling $\hat{\mathbf{x}}$.

5.1. Overestimation of the Energy Function

Given an interval $\mathbf{I} = \{f, \cdots, l\}$ of consecutive labels, and the current labeling $\hat{\mathbf{x}}$, we define the new energy function $E'(\cdot)$ over the set of random variables $\mathbf{X}$. Unlike the original energy function, the label set corresponding to $E'(\cdot)$ is equal to $\mathbf{L}' = \{0, 1, \cdots, h'\}$, where $h' = l - f + 1$. The label 0 in the set $\mathbf{L}'$ corresponds to a random variable retaining its current label, while any other label $i \geq 1$ corresponds to a random variables taking the label $f + i - 1 \in \mathbf{I}$. A labeling of the energy function $E'(\cdot)$ is denoted by $\mathbf{y} \in (\mathbf{L}')^n$ in order to distinguish it from the labeling corresponding to the original energy function. We say that a labeling $\mathbf{x} \in \mathbf{L}^n$ corresponds to the labeling $\mathbf{y} \in (\mathbf{L}')^n$ if

$$x_a = \begin{cases} \hat{x}_a & \text{if } y_a = 0, \\ y_a + f - 1 & \text{otherwise.} \end{cases} \tag{4}$$

We define the unary potentials and the clique potentials of the energy function $E'(\cdot)$ as follows.

Unary Potentials. The unary potential of a random variable X_a (where $a \in \mathcal{V}$) being assigned a label $y_a \in \mathbf{L}'$ is given by the following equation:

$$\theta'_a(y_a) = \begin{cases} \theta_a(\hat{x}_a) + \kappa_a & \text{if } y_a = 0 \\ \theta_a(y_a + f - 1) + \kappa_a & \text{otherwise.} \end{cases} \tag{5}$$

In other words, if $y_a = 0$ then the unary potential corresponds to the random variable X_a retaining its current label

$\hat{x}_a$, and if $y_a \neq 0$ then the unary potential corresponds to the random variable X_a being assigned the label $y_a + f - 1 \in \mathbf{I}$. The constant κ_a is added to the unary potentials to ensure that they are non-negative, which makes the description of the graph construction in the next subsection simpler.

Clique Potentials. In order to describe the high-order clique potentials of the new energy function we require a function $\delta_{a,b} : \mathbf{L}' \times \mathbf{L}' \rightarrow \mathbb{R}$ for each $(a, b) \in \mathcal{E}$, which is defined as follows:

$$\delta_{a,b}(y_a, y_b) = \begin{cases} \min\{d(\hat{x}_a - \hat{x}_b), M\} & \text{if } y_a = y_b = 0, \\ M + d(y_b - 1) & \text{if } y_a = 0, y_b \neq 0, \\ M + d(y_a - 1) & \text{if } y_a \neq 0, y_b = 0, \\ d(y_a - y_b) & \text{if } y_a \neq 0, y_b \neq 0. \end{cases} \tag{6}$$

Here, $d(\cdot)$ is the convex function and M is the truncation factor associated with the original energy function $E(\cdot)$.

Proposition 1 *$\delta_{a,b}(y_a, y_b)$ is submodular in the sense of label-set encoding used in [13] and is an overestimate of the truncated convex distance* $\min\{d(y_a - y_b), M\}$.

The proof of proposition 1 can be found in the supplementary. Note that this submodular upper bound is not tight and has been chosen to facilitate the analysis of multiplicative bounds later. Given a labeling $\mathbf{y_c} \in (\mathbf{L}')^c$ of a clique $\mathbf{c}$ of size c, we denote a sorted list of the labels in $\mathbf{y_c}$ as $\mathbf{p}(\mathbf{y_c})$. Furthermore, we denote the indices of the sorted list as $\mathbf{q}(\mathbf{y_c})$. In other words, the random variable corresponding to the i-th smallest label (that is, the i-th element of the list $\mathbf{p}(\mathbf{y_c})$, which is denoted by $p_i(\mathbf{y_c})$) is given by X_a where $a = q_i(\mathbf{y_c})$. To avoid clutter, we will drop the argument $\mathbf{y_c}$ from $\mathbf{p}$ and $\mathbf{q}$ whenever it is clear from context.

Using the above definitions, the high-order clique potential for the new energy $E'(\cdot)$ can be concisely specified as

$$\theta'_{\mathbf{c}}(\mathbf{y_c}) = \omega_{\mathbf{c}} \sum_{i=1}^{m} \delta_{q_i, q_{c-i+1}}(p_i, p_{c-i+1}). \tag{7}$$

Hence, the clique potentials in the energy function $E'(\cdot)$ are the sum of the m maximum submodular functions over disjoint pairs of random variables in the clique.

5.2. Graph Construction

Our problem is to minimize the energy function $E'(\cdot)$ over all possible labelings $\mathbf{y} \in (\mathbf{L}')^n$. To this end, we convert it into an equivalent st-MINCUT problem over a directed graph, which can be solved efficiently if all arc capacities are non-negative [2].

We construct a directed graph over the set of vertices $\{s, t\} \cup \mathbf{V} \cup \mathbf{U} \cup \mathbf{W}$. The set of vertices $\mathbf{V}$ model the random variables $\mathbf{X}$. Specifically, for each random variable X_a we define $h' = l - f + 1$ vertices V_i^a where $i \in \{1, \cdots, h'\}$. The sets $\mathbf{U}$ and $\mathbf{W}$ represent *auxiliary* vertices, whose role in the graph construction will be explained later when we consider representing the high-order clique potentials. We also define a set of arcs over the vertices, where each arc has a non-negative capacity. We would like to assign arc capacities such that the st-cuts of the directed graph satisfy two properties. First, all the st-cuts with a finite capacity should include exactly one arc from the set $(s, V_1^a) \cup \{(V_i^a, V_{i+1}^a), i = 1, \cdots, h' - 1\} \cup (V_{h'}^a, t)$ for each random variable X_a. This property would allow us to define a labeling $\mathbf{y}$ such that

$$y_a = \begin{cases} 0 & \text{if the cut includes the arc } (s, V_1^a) \\ i & \text{if the cut includes the arc } (V_i^a, V_{i+1}^a) \\ h' & \text{if the cut includes the arc } (V_{h'}^a, t). \end{cases} \tag{8}$$

Second, we would like the energy of the labeling $\mathbf{y}$ defined above to be as close as possible to the capacity of the st-cut. This will allow us to obtain an optimal labeling with respect to the energy function $E'()$ by finding the st-MINCUT. We now specify the arcs and their capacities such that they satisfy the above two properties. We consider two cases: (i) arcs that represent the unary potentials; and (ii) arcs that represent the high-order clique potentials.

Representing Unary Potentials. We will represent the unary potential of X_a using the arcs specified in Figure 2. Since all the unary potentials are non-negative, it follows that the arc capacities in Figure 2 are also non-negative.

Representing Clique Potentials. Consider a set of random variables $\mathbf{X_c}$ that are used to define a high-order clique potential. Without loss of generality, we assume $\mathbf{X_c} = \{X_1, X_2, \cdots, X_c\}$. In order to represent the potential value for a putative labeling $\mathbf{x_c}$ of the clique, we introduce two types of arcs, which are depicted in Figure 3. For the arcs shown in Figure 3 (left), the capacities are specified using the term r_{ij} that is defined as follows:

$$r_{ij} = \begin{cases} \omega_{\mathbf{c}} \frac{\overline{d}(i,j)}{2} & \text{if } i = j \neq 1 \\ \omega_{\mathbf{c}} \overline{d}(i, j) & \text{if } i > j \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Here, the term $\overline{d}(i, j) = d(i - j + 1) + d(i - j - 1) - 2d(i - j) \geq 0$ since $d(\cdot)$ is convex, and $\omega_{\mathbf{c}} \geq 0$ by definition. It follows that $r_{ij} \geq 0$ for all $i, j \in \{1, \cdots, h'\}$. For the arcs shown in Figure 3 (right), the capacities are specified using the terms A and B that are defined as follows:

$$A = \omega_{\mathbf{c}} M, B = \left(\omega_{\mathbf{c}} M - \frac{\theta_{\mathbf{c}}(\hat{\mathbf{x}}_{\mathbf{c}})}{m}\right). \tag{10}$$

Since $M \geq 0$, and $\theta_{\mathbf{c}}(\hat{\mathbf{x}}_{\mathbf{c}}) \leq \omega_{\mathbf{c}} m M$ due to truncation, it follows that $A, B \geq 0$.

While it is not immediately obvious that the above arcs allow us to represent the clique potential values as the capacities of the correspondings cuts, the following proposition establishes this desired property.

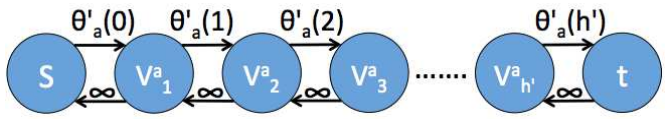


Figure 2: *Arcs and their capacities for representing the unary potentials for the random variable X_a. According to the labeling defined in equation (8), if $x_a = \hat{x}_a$, then the arc (s, V_1^a) will be cut, which will contribute exactly $\theta_a(\hat{x}_a)$ to the capacity of the cut. If $x_a = s + i - 1$ where $i \in \{1, \cdots, h' - 1\}$, then the arc (V_i^a, V_{i+1}^a) will be cut, which will contribute exactly $\theta_a(s + i - 1)$ to the capacity of the cut. If $x_a = l$, then the arc $(V_{h'}^a, t)$ will be cut, which will contribute exactly $\theta_a(l)$ to the capacity of the cut. The arcs with infinite capacity ensure that exactly one of the arcs from the set $(s, V_1^a) \cup \{(V_i^a, V_{i+1}^a), i = 1, \cdots, h' - 1\} \cup (V_{h'}^a, t)$ is part of an st-cut with finite capacity, which guarantees that we are able to obtain a valid labeling.*

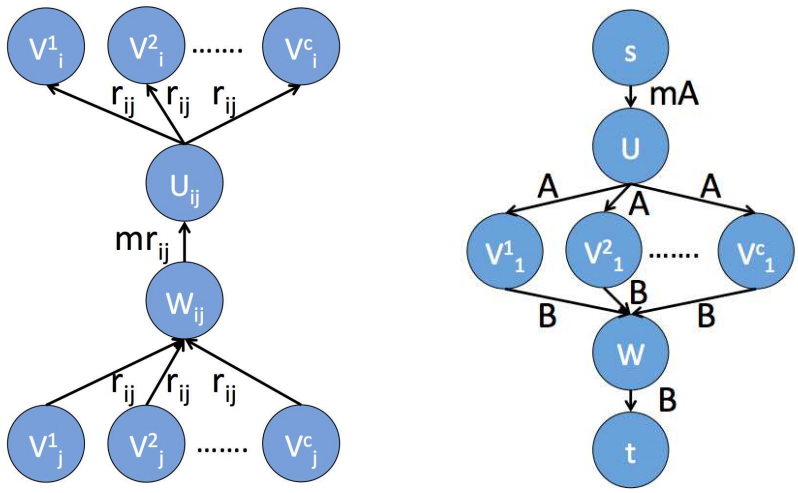


Figure 3: *Arcs used to represent the high-order potentials for the clique $\mathbf{X_c} = \{X_1, X_2, \cdots, X_c\}$.* **Left.** *The term r_{ij} is defined in equation (9). The arcs represent the sum of the m maximum convex distance functions over disjoint pairs of random variables when no random variable retains its old label. These arcs are specified only for $i \leq j$ and when either one or both of i and j are not equal to 1.* **Right.** *The terms A and B are defined in equation (10). The arcs represent an overestimation of the clique potential for the case where some or all the random variables retain their old label.*

Proposition 2 *Given a cut that partitions the vertices $\mathbf{V}$ into two disjoint sets $\mathbf{V}_s$ and $\mathbf{V}_t$, and the corresponding labeling $\mathbf{y}$ as defined in equation (8), the capacity of the cut is equal to the energy $E'(\mathbf{y})$ up to a constant.*

The proof of the above proposition is provided in the supplementary material. The following corollary may be of interest to the reader as it shows that our graph construction generalizes that of [13] for pairwise convex potentials.

Corollary 1 *The above graph construction can be used to exactly minimize an energy function that consists of arbitrary unary potentials and clique potentials that are proportional to the sum of m maximum convex distances over all disjoint pairs of random variables present in the clique.*

Energy Minimization. The above proposition implies that the labeling $\mathbf{y}'$ corresponding to the st-MINCUT minimizes the energy $E'(\cdot)$ over all possible labelings $\mathbf{y} \in (\mathbf{L}')^n$. Since all the arc capacities in the graph are non-negative, the labeling $\mathbf{y}'$ can be computed efficiently by solving the st-MINCUT problem on the directed graph defined above. Once the labeling $\mathbf{y}'$ is computed, we find an approximate solution $\mathbf{x}'$ to problem (3) using equation (4).

5.3. Multiplicative Bounds

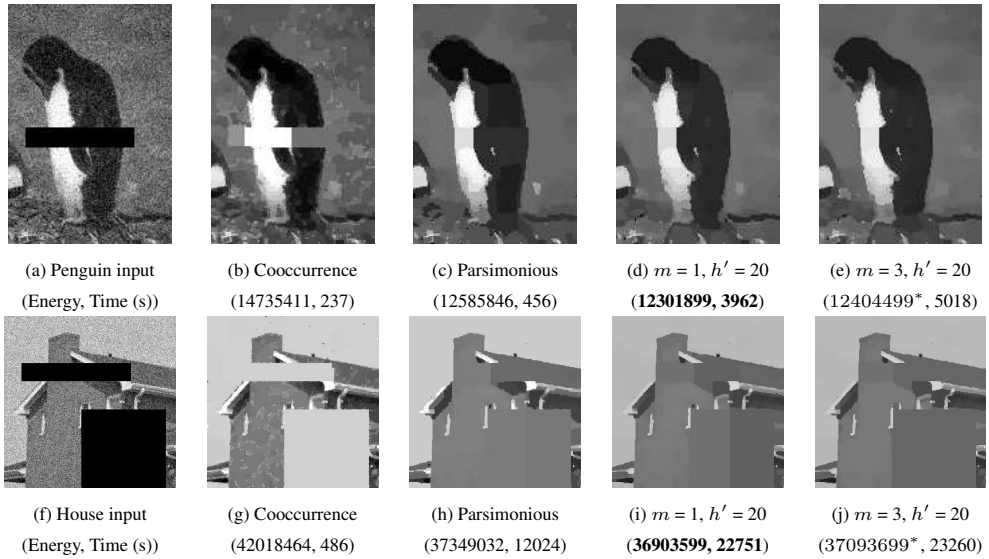
We conclude our technical description by establishing its strong theoretical guarantees for special cases of interest. We give multiplicative bound on our final solutions, which also serve to identify the best value of the interval length parameter h'.

Proposition 3 *The range expansion algorithm with $h' = M$ has a multiplicative bound of $O(m \cdot C)$ for truncated max-of-linear model for any general value of m. The term C equals the size of the largest clique. Hence, if $\mathbf{x}^*$ is a labeling with minimum energy and $\hat{\mathbf{x}}$ is the labeling estimated by range expansion algorithm then*

$$\sum_{a \in \mathcal{V}} \theta_a(\hat{x}_a) + \sum_{\mathbf{c} \in \mathcal{C}} \theta_{\mathbf{c}}(\hat{\mathbf{x}}_{\mathbf{c}}) \leq \sum_{a \in \mathcal{V}} \theta_a(x_a^*) + O(m \cdot C) \sum_{\mathbf{c} \in \mathcal{C}} \theta_{\mathbf{c}}(\mathbf{x}_{\mathbf{c}}^*).$$

The above inequality holds for arbitrary set of unary potentials and non-negative clique weights.

Note that for $m = 1$, the bound of the move making algorithm for parsimonious labeling (baseline) is $\left(\frac{r}{r-1}\right)(C - 1) \cdot \min(C, h) \cdot O(\log h)$ where C is the size of the largest clique, h is the number of labels, and r is a parameter in the algorithm [8]. Our algorithm gives a better bound of $O(C)$ and does not depend on the number of labels.



(a) Penguin input (Energy, Time (s)) | (b) Cooccurrence (14735411, 237) | (c) Parsimonious (12585846, 456) | (d) m = 1, h' = 20 (**12301899, 3962**) | (e) m = 3, h' = 20 (12404499*, 5018)

(f) House input (Energy, Time (s)) | (g) Cooccurrence (42018464, 486) | (h) Parsimonious (37349032, 12024) | (i) m = 1, h' = 20 (**36903599, 22751**) | (j) m = 3, h' = 20 (37093699*, 23260)

Figure 4: *Image inpainting results: Figures (a) and (f) are input images of 'penguin' and 'house' respectively with noise and obscured regions. Our results for m = 1 (d) and (i) are significantly better than those of [25] (b) and (g) and of [8] (c) and (h) in terms of energy. We use super-pixels obtained using mean-shift as cliques. Our results preserve details better and look more natural. The baseline results exhibit significant blocking effect. *Note that m = 3 uses a different energy function from other cases.*

Proposition 4 *The range expansion algorithm with $h' = \sqrt{M}$ has a multiplicative bound of $O(C\sqrt{M})$ for the truncated max-of-quadratic model when $m = 1$.*

The proofs for the above two propositions are provided in the supplementary material. The proofs for these proceed by establishing a lower bound on the energy improvement at each iteration. This lower bound is obtained by the energy difference between the current labeling and a new labeling obtained by assigning all random variables either their optimum labels if the optimum labels lie in the current interval, else retaining their old labels. The energy of the latter labeling is an upper bound on the true energy of the new labeling obtained by solving the st-MINCUT problem.

6. Experiments

To demonstrate the efficacy of our algorithm, we test it on the two problems of image inpainting and denoising, and stereo matching. The final labeling energy and convergence time are used as evaluation criteria. We used the parsimonious labeling algorithm of Dokania *et al.* [8] and the move-making algorithm for the co-occurrence based energy function of Ladicky *et al.* [25] as baselines. For comparison, we restrict ourselves to max-of-linear models and m = 1, as the available code for the baselines can only handle this special case. For completeness, we report the results of our range expansion algorithm for other cases of TMCM and on synthetic data (in supplementary) as well. Note that the energy values for different m are not comparable since they make use of different models.

6.1. Image Inpainting and Denoising

Data. Given an image with noise and obscured/inpainted regions (regions with missing pixels), the task is to denoise it and fill the obscured regions in a way that is consistent with the surrounding regions. The images 'house' and 'penguin' from the Middlebury data set were used for the experiments. Since the images are grayscale, they have 256 labels in the interval [0, 255], each representing an intensity value. The unary potential for each pixel corresponding to a particular label equals the squared difference between the label and the intensity value at the pixel. We use high-order cliques as the super-pixels obtained using the mean-shift method [5]. The parameters ω_c, M and m are varied to give different truncated max-of-linear energy functions.

Method. For each parameter setting of ω_c, M and m, we vary h' for our algorithm and compare with the baselines.

Results. Results for ω_c = 50, M = 50, and m = 1 and 3 for 'house' and ω_c = 40, M = 40, and m = 1 and 3 for 'penguin' are shown in Figure 4. Other results are shown in the supplementary material. In both cases, we used interval length h' = 20. The value of h' giving best results in practice differs from that suggested in Proposition 3 for optimal bound. Our algorithm consistently gives lower energy labeling as compared to both [8] and [25]. For 'penguin', our algorithm gives cleaner denoised image, preserving edges and details. On the other hand, both [8] and [25] exhibit significant blocking effect. Moreover, the output is more natural for m = 3 as compared to m = 1. Even for 'house', our output looks more visually appealing than the baselines.

6.2. Stereo Matching

Data. In the stereo matching problem, we have two rectified images of the same scene from two cameras set slightly apart. We need to estimate the horizontal disparity between a pixel in the right camera image from the corresponding

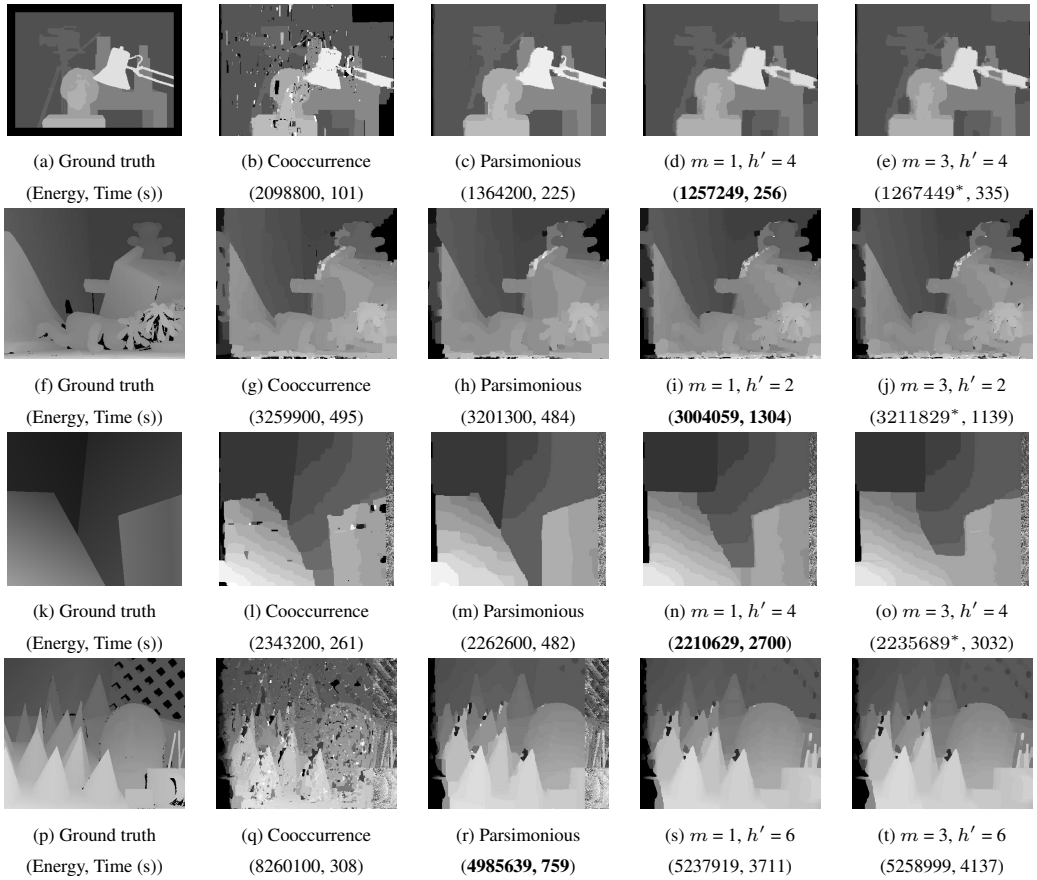


Figure 5: *Stereo matching results: Figures (a), (f) (k) and (p) are ground truth disparity for 'tsukuba', 'teddy', 'venus' and 'cones' respectively. Apart from 'cones', our results for $m = 1$ (d), (i) and (n) are significantly better than those of [25] (b), (g) and (l) and of [8] (c), (h) and (m) in terms of energy. We also show results for $m = 3$. We use super-pixels obtained using mean-shift as cliques. *Note that $m = 3$ uses a different energy function from other cases.*

pixel in the left camera. We use 'tsukuba' and 'teddy' data sets from the Middlebury stereo collection for our experiments. In each case, we have a pair of RGB images and ground truth disparities. We assume the unary potentials to be the $L1$-norm of the difference in RGB values of the left and right image pixels. There are 16 labels for 'tsukuba', 20 for 'venus', and 60 for 'teddy' and 'cones'. The high-order cliques are super-pixels obtained using mean-shift method [5]. The parameters ω_c, M and m are varied to give different truncated max-of-linear energy functions.

Method. For each parameter setting of ω_c, M and m, we vary h' for our algorithm and compare with the baselines.

Results. Results for $\omega_c = 20$, $M = 5$, and $m = 1$ and 3 for 'tsukuba', 'venus' and 'cones' and $\omega_c = 20$, $M = 1$, and $m = 1$ and 3 for 'teddy' are shown in Figure 5. We used interval length h' as 4 for 'tsukuba' and 'venus', 6 for 'cones' and 2 for 'teddy'. Apart from 'cones', our algorithm consistently gives lower energy labeling as compared to both [8] and [25]. For 'tsukuba', our algorithm captures the details of the face better than [8] and [25]. For 'venus', we get smoother labeling for the front plane. Moreover, our results for $m = 3$ exhibit robustness to inaccurate clique definitions.

7. Discussion

We proposed a novel family of high-order random fields called truncated max-of-convex models (TMCM) which are generalization of truncated convex models (TCM). To perform inference in TMCM, we developed a novel range expansion algorithm for energy minimization that retains the efficiency of st-MINCUT and provides provably accurate solutions. The algorithm relies on an exact graph representation of max-of-convex models, a submodular overestimate of the energy function for any interval length and a graph construction that represents this overestimate, allowing the inference problem to be solved using st-MINCUT. Theoretically, our work can be thought of as a step towards the identification of graph representable submodular functions and automated construction of graphs for such functions.

8. Acknowledgement

This work was supported by the Google DeepMind PhD Studentship and by the Alan Turing Institute under the EPSRC grant EP/N510129/1. We thank Rudy Bunel, Diane Bouchacourt, and Stuart Golodetz for their useful discussions and insights.

References

[1] C. Arora and S. Maheshwari. Multi label generic cuts: Optimal infernce in multi label multi clique MRF-MAP problems. In *CVPR*, 2014. 2

[2] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI*, 2004. 2, 5

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 2001. 1, 2, 3

[4] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. Approximation algorithms for the metric labeling problem via a new linear programming formulation. In *SODA*, 2001. 1, 3

[5] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 2002. 7, 8

[6] A. Delong, L. Gorelick, O. Veksler, and Y. Boykov. Minimizing energies with hierarchical costs. *IJCV*, 2012. 2

[7] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. In *CVPR*, 2010. 2

[8] P. Dokania and M. P. Kumar. Parsimonious labeling. In *ICCV*, 2015. 2, 6, 7, 8

[9] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 2006. 2

[10] A. Fix, C. Wang, and R. Zabih. A primal-dual algorithm for higher-order multilabel Markov random fields. In *CVPR*, 2014. 2

[11] S. Gould, F. Amat, and D. Koller. Alphabet soup: A framework for approximate energy minimization. In *CVPR*, 2009. 2

[12] A. Gupta and E. Tardos. A constant factor approximation algorithm for a class of classification problems. In *STOC*, 2000. 1, 3

[13] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *PAMI*, 2003. 5, 6

[14] S. Jegelka and J. Bilmes. Submodularity beyond submodular energies: coupling edges in graph cuts. In *CVPR*, 2011. 2

[15] P. Kohli, M. P. Kumar, and P. Torr. P3 & beyond: Solving energies with higher order cliques. In *CVPR*, 2007. 2, 3

[16] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008. 2, 3

[17] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 2006. 1, 3

[18] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 2006. 2

[19] V. Kolmogorov. Minimizing a sum of submodular functions. *Discrete Applied Mathematics*, 2012. 2

[20] N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *ICCV*, 2007. 1, 3

[21] N. Komodakis, N. Paragios, and G. Tziritas. Mrf optimization via dual decomposition: Message-passing revisited. In *ICCV*, 2007. 2

[22] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic MRFs. In *CVPR*, 2007. 1, 3

[23] M. P. Kumar and P. Torr. Improved moves for truncated convex models. In *NIPS*, 2008. 3

[24] M. P. Kumar, O. Veksler, and P. H. Torr. Improved moves for truncated convex models. *JMLR*, 2011. 1

[25] L. Ladicky, C. Russell, P. Kohli, and P. Torr. Graph cut based inference with co-occurrence statistics. In *ECCV*, 2010. 2, 7, 8

[26] O. Meshi, D. Sontag, T. S. Jaakkola, and A. Globerson. Learning efficiently with approximate inference via dual losses. In *ICML*, 2010. 2

[27] D. Sontag, T. Meltzer, A. Globerson, T. S. Jaakkola, and Y. Weiss. Tightening LP relaxations for map using message passing. In *UAI*, 2008. 1

[28] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *PAMI*, 2008. 1, 2

[29] D. Tarlow, I. E. Givoni, and R. S. Zemel. Hop-map: Efficient message passing with high order potentials. In *AISTATS*, 2010. 2

[30] O. Veksler. Graph cut based optimization for MRFs with truncated convex priors. In *CVPR*, 2007. 1, 3

[31] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. MAP estimation via agreement on trees: message-passing and linear programming. *IEEE Trans. Information Theory*, 2005. 2