

Truncated Max-of-Convex Models

Pankaj Pansari

Department of Engineering Science
University of Oxford
pankaj@robots.ox.ac.uk

M. Pawan Kumar

Department of Engineering Science
University of Oxford
pawan@robots.ox.ac.uk

Abstract

Truncated convex models (TCM) are special cases 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 potentials, which have no restriction on their form; and (ii) high-order potentials, which are the sum of the truncation of the m largest convex distances over disjoint pairs of random variables in an arbitrary size clique. The use of a convex distance function encourages smoothness, while truncation allows for discontinuities in the labeling. By using $m > 1$, TMCM provides robustness towards errors in the clique definition. In order 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.

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. For example, in image denoising, each random variable could correspond to the true unknown intensity value of a pixel, while the label set could be the putative intensity values $\{0, \dots, 255\}$. The neighborhood relationship could correspond to a 4 or 8 neighborhood.

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 two types of potentials. First, a unary potential that depends on the label of one ran-

dom variable, and has no restriction on its form. Second, a pairwise potential that depends on the labels of two neighboring random variables. The pairwise potential is proportional to the truncated convex distance between the labels. The use of a convex distance function encourages smoothness in the labeling, while the use of truncation allows for discontinuities in the labeling.

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, 11, 15, 17, 18, 19, 23], which provide accurate solutions in practice [22]. In fact, there are compelling theoretical reasons to believe that the linear programming relaxation based approaches [4, 15, 19] are the best polynomial-time algorithms that can be devised for energy minimization [21]). Of particular interest to us is the range expansion algorithm that combines the efficiency of st -MINCUT with the accuracy of linear programming [19].

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 truncation of the m largest convex distance functions computed over disjoint pairs of random variables in the clique. Here, disjoint pairs refer to the fact that the label of no random variable is used more than once to compute the value of the clique potential. The term m is a positive integer that is less than or equal to half the clique size. Note that the use of disjoint pairs is not restrictive as we can introduce

dummy random variables in a clique that are forced to take the same label as another random variable. 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, instead of using pairwise potentials that measure the difference in intensity of two neighboring pixels, we can use the variance of intensity values over a superpixel.

In order to enable the use of TMCs 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 grow linearly with the number of random variables, the number of labels and the number of cliques). 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.

2. Related Work

Pairwise truncated convex models (TCM) offer a natural framework to capture low-level cues for vision problems such as image denoising, stereo correspondence, segmentation and optical flow [22]. Specifically, through the use of convex distance functions they encourage smoothness in the labeling. Smooth labelings are desirable in low-level vision since images typically have large homogeneous regions that correspond to a single object. At the same time, TCM allow for discontinuities in the labeling, which is expected to occur at edge pixels between two objects. Their use is also supported by the availability of a vast number of highly efficient and accurate energy minimization algorithms [3, 4, 11, 15, 17, 18, 19, 23]. However, the restriction to pairwise potentials limits their representational power.

For the past few years, the computer vision community has witnessed a growing interest in high-order models. 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 of the earliest such works is the P^n Potts model [13], which encourages label consistency over a set of random variables. This work has extended in [14], 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 [10], 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.

Delong et al. [5, 6] 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 [7] 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 truncated convex models. The only metric that also results in a truncated convex model is the truncated linear distance.

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. [8] use the submodular max-flow algorithm [16], 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. [20], who proposed a co-occurrence based clique potential whose only requirement is that it should increase monotonically with the set of unique labels represent in the clique labeling. However, the use of such a general clique potential still results in an inaccurate energy minimization algorithm.

3. Truncated Convex Models

Before describing our model in detail, we briefly review the standard pairwise TCM, which will help set up the necessary notation. 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 so as to enable the use of convex distance functions. Without loss of generality, we define $\mathcal{V} = \{1, 2, \dots, n\}$ and $\mathbf{L} = \{1, 2, \dots, h\}$.

An assignment of values to all the random variables $\mathbf{x} \in \mathbf{L}^n$ is referred to as a labeling. In order to quantitatively distinguish the h^n possible labelings of the random variables, a TCM defines an energy function that consists of two types of potentials. First, the unary potentials $\theta_a(x_a)$ that depend on the label x_a of one random variable X_a . Second, the pairwise potentials $\theta_{ab}(x_a, x_b)$ that depend 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.

In order 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 $y, z \in \mathbb{Z}$. Popular examples of convex distance functions include the linear distance (that is, $d(y) = |y|$) and the quadratic distance function (that is, $d(y) = y^2$).

Given two labels $i, j \in \mathbf{L}$, we can use a convex function $d(\cdot)$ to measure the distance between them as $d(i - j)$. By specifying a pairwise potential that is proportional to the convex distance between the labels assigned to the neighboring random variables, we can encourage *smooth* labelings as they will correspond to a lower energy value. The ability of convex distance functions to encourage smoothness makes them highly suited for low-level vision problems as images tend to consist of large homogenous regions. However, images also naturally contain some discontinuities (for example, the intensity values of edge pixels differ greatly). In order to prevent the overpenalization of the discontinuities, it is common practice to use a truncated convex distance function over the label set [3, 19, 23]. Formally, a truncated convex function is defined as $\min\{d(\cdot), M\}$, where M is the truncation factor. Using a truncated convex function, we 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.

To summarize, a TCM specifies an energy function

$E(\cdot)$ over the labelings $\mathbf{x} \in \mathbf{L}^n$ as follows:

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

The unary potentials are arbitrary, the edge weights ω_{ab} are non-negative, $d(\cdot)$ is a convex function and $M \geq 0$ is the truncation factor. 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 [19], as well as several other approaches based on *st*-MINCUT [3, 11, 18, 23] and linear programming [4, 15, 17].

4. Truncated Max-of-Convex Models

Our aim is to obtain 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 want to place any restriction on the size of the clique. The main difficulty is a lack of a natural distance measure over a label set of arbitrary size. Recall that, in the pairwise case, we did not face this problem as the truncated convex function could be used to measure the distance between the two labels assigned to the neighboring random variables. To alleviate this difficulty, we propose to use the *sum of the truncation of the m largest convex distances over disjoint pairs of random variables* that belong to the clique. We first formally specify the form of our high-order clique potentials. This will allow us to subsequently discuss their advantages in modeling low-level vision applications.

Truncated Max-of-Convex Potentials. Consider a high-order clique consisting of the random variables $\mathbf{X}_{\mathbf{c}} = \{X_a, a \in \mathbf{c} \subseteq \mathcal{V}\}$. We denote a labeling of the clique as $\mathbf{x}_{\mathbf{c}} \in \mathbf{L}^c$, 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 $\mathbf{x}_{\mathbf{c}}$ we require a sorted list of the (not necessarily unique) labels present in $\mathbf{x}_{\mathbf{c}}$. We denote this sorted list by $\mathbf{p}(\mathbf{x}_{\mathbf{c}})$ and access its i -th element as $p_i(\mathbf{x}_{\mathbf{c}})$. For example, consider a clique consisting of random variables $\mathbf{X}_{\mathbf{c}} = \{X_1, X_2, X_3, X_4, X_5, X_6\}$. If the number of labels $h = 10$, then one of the putative labelings of the clique is $\mathbf{x}_{\mathbf{c}} = \{3, 2, 1, 4, 1, 3\}$ (that is, X_1 takes the value 3, X_2 takes the value 2 and so on). For this labeling, $\mathbf{p}(\mathbf{x}_{\mathbf{c}}) = \{1, 1, 2, 3, 3, 4\}$. The value of $p_1(\mathbf{x}_{\mathbf{c}})$ and $p_2(\mathbf{x}_{\mathbf{c}})$ is 1, the value of $p_3(\mathbf{x}_{\mathbf{c}})$ 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

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

Here, $\omega_{\mathbf{c}} \geq 0$ is the clique weight that does not depend on the labeling. However, it can depend on the data corresponding to the random variables in the clique, and is therefore capable of modeling the statistics of a set of pixels. This is in contrast to the pairwise truncated convex models that normally rely on the difference in appearance of only two neighboring pixels. 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 mentioned earlier, the use of disjoint pairs in the definition of our potentials is not restrictive as we can always introduce dummy random variables that are forced to take the same label as another random variable in the clique via appropriate pairwise potentials.

In the remaining part of this section, we will analyze the advantages of our high-order potentials from a modeling point of view. The next section will demonstrate their computational advantage by generalizing the efficient range expansion algorithm to accurately minimize energy functions that consist of arbitrary unary potentials and high-order truncated max-of-convex potentials.

Labeling	$m = 1$	$m = 2$	$m = 3$
$\{1,1,1,1,2,2\}$	1	2	2
$\{1,2,3,4,5,6\}$	3	6	7
$\{1,1,1,9,9,9\}$	3	6	9
$\{1,1,1,8,8,9\}$	3	6	9
$\{1,1,1,1,1,7\}$	6	6	6
$\{1,1,1,2,3,4\}$	3	5	6

Table 1. The value of the clique potential defined by a linear function with a truncation factor $M = 3$ for various values of the parameter m . Since the size of the clique is 6, $0 \leq m \leq 3$. The first pair of labelings demonstrates why taking the largest convex distances favors smoothness. The next pair demonstrates how the truncation prevents overpenalizing discontinuities. The last pair demonstrates how the use of $m > 1$ can provide some degree of robustness to errors in the definitions of the cliques.

Smoothness. The truncated max-of-convex potentials encourage smooth labelings. In order to illustrate

this, let us consider the example of a clique of six random variables $\mathbf{X}_{\mathbf{c}}$ and a label set \mathbf{L} of size 10. We can define a truncated convex distance using a linear function $d(y) = |y|$ and a truncation factor of $M = 3$. Consider the first pair of labelings shown in table 1. Clearly, the first labeling of this pair is significantly smoother than the second, which is reflected in the value of the clique potential for all values of m . In contrast, if we were to consider the minimum distance among all pairs of labels, both the labelings will provide a clique potential value of 0.

Discontinuities. Similar to the pairwise case, the use of a truncation factor helps prevent the overpenalization of discontinuities. For example, let us consider the second pair of labelings in table 1. In both the cases, the six random variables appear to belong to two *groups*, one whose labels are low and one whose labels are high. Without a truncation, such a discontinuity would have been penalized heavily (for example, 8 for $m = 1$ for both the labelings). This in turn would discourage the clique to be assigned this labeling even though this type of discontinuity is expected to occur in natural images. However, with a truncation, the penalty is significantly less (for example, 3 for $m = 1$ for both the labelings), which can help preserve the discontinuities in the labeling obtained via energy minimization.

Robustness. In order to use a TMCM, we would be required to define the cliques. For example, given an image, we could use a bottom-up oversegmentation approach to obtain superpixels, and then set all the pixels in a superpixel to belong to a clique. However, oversegmentation can introduce errors since it has no notion of the specific vision application we are interested in modeling. To add robustness to errors in the clique definitions, we can set $m > 1$. For example, consider the final pair of labelings in table 1. The first of these labelings contains a single random variable with a very high label, which could be due to the fact that the corresponding pixel has been incorrectly grouped in this superpixel. As can be seen from the values of the potential, the presence of such an erroneous pixel in the superpixel is not heavily penalized when we use $m > 1$. For example, when $m = 3$ the value of the clique potential for the first labeling (with an erroneous pixel) is the same as the second labeling (which is a fairly smooth labeling).

To summarize, a TMCM specifies the following energy function $E(\cdot)$ over the labelings $\mathbf{x} \in \mathbf{L}^n$:

$$E(\mathbf{x}) = \sum_{a \in \mathcal{V}} \theta_a(x_a) + \sum_{\mathbf{c} \in \mathcal{C}} \theta_{\mathbf{c}}(\mathbf{x}_{\mathbf{c}}). \quad (3)$$

Here, the unary potentials $\theta_a(\cdot)$ are arbitrary, the clique potentials $\theta_c(\cdot)$ are as defined in equation (2), and \mathcal{C} refers to the set of all cliques in the random field. Given an input, the desired output is obtained by solving the following optimization problem: $\min_{\mathbf{x} \in \mathcal{L}^n} E(\mathbf{x})$. Henceforth, we assume the unary potentials are non-negative. Note that 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 will be seen shortly, our algorithm as well as its theoretical guarantees are invariant to such changes in the energy function.

5. Optimization via Range Expansion

As TMCs are 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. Due to space limitations, we only provide an overview of our approach. The algorithm is described in detail in the supplementary material, which also contains the proofs of our propositions.

The range expansion algorithm is an iterative move-making method, which starts with an initial labeling \mathbf{x}^0 . At each iteration, given the current labeling $\hat{\mathbf{x}}$ it attempts to move to a better labeling by allowing each random variable X_a to either retain its old label \hat{x}_a or choose a new label from an interval $\mathbf{I} = \{s, \dots, l\}$ of length $h' = l - s + 1$. In other words, a new labeling is chosen by solving the following optimization problem:

$$\begin{aligned} \mathbf{x}' = & \underset{\mathbf{x}}{\operatorname{argmin}} E(\mathbf{x}), \\ \text{s.t. } & x_a \in \mathbf{I} \cup \{\hat{x}_a\}, \forall a \in \mathcal{V}. \end{aligned} \quad (4)$$

At each iteration, the range expansion algorithm chooses a new interval \mathbf{I} of length h' . The algorithm converges when the energy of the labeling cannot be reduced further for any choice of the interval.

The main challenge we face is that problem (4) itself may be NP-hard. Indeed, when $h' = h$, problem (4) is equivalent to the original energy minimization problem. To alleviate this difficulty, we propose an *st*-MINCUT based approximate solution to the above problem. In other words, we construct a directed graph whose cuts encode the putative labelings of problem (4). We obtain an approximate solution \mathbf{x}' by computing the *st*-MINCUT. The graph construction for a given current labeling $\hat{\mathbf{x}}$ and an interval \mathbf{I} of some arbitrary length h' is described in the next subsection. Subsection 5.2 provides strong theoretical guarantees for various special cases of interest. This serves two

purposes: (i) it establishes the accuracy of range expansion for TCM; and (ii) it helps identify the optimal value of the interval length parameter.

5.1. Graph Construction

We would like to minimize the energy function $E(\cdot)$ over all possible labelings that allow each random variable X_a to either retain its current label \hat{x}_a or choose a label from the interval $\mathbf{I} = \{s, \dots, l\}$. To this end, we convert it into an equivalent *st*-MINCUT problem over a directed graph. Recall that an *st*-cut partitions the vertices of a directed graph into two disjoint subsets \mathbf{V}_s and \mathbf{V}_t such that the source vertex $s \in \mathbf{V}_s$ and the sink vertex $t \in \mathbf{V}_t$. The capacity of the cut is the sum of the capacity of all the arcs whose starting vertex is in \mathbf{V}_s and whose ending vertex is in \mathbf{V}_t . In other words, only the arcs going from \mathbf{V}_s to \mathbf{V}_t contribute to the capacity of an *st*-cut, and not the arcs going from \mathbf{V}_t to \mathbf{V}_s . The *st*-MINCUT problem, that is, finding an *st*-cut with the minimum capacity, 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 $h' = l - s + 1$ vertices V_i^a where $i \in \{1, \dots, h'\}$. 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 $(s, V_1^a) \cup \{(V_i^a, V_{i+1}^a), i = 1, \dots, h' - 1\} \cup (V_{h'}^a, t)$ for each random variable X_a . This property would allow us to define a labeling \mathbf{x} such that

$$x_a = \begin{cases} \hat{x}_a & \text{if the cut includes the arc } (s, V_1^a) \\ s + i - 1 & \text{if the cut includes the arc } (V_i^a, V_{i+1}^a) \\ l & \text{if the cut includes the arc } (V_{h'}^a, t). \end{cases} \quad (5)$$

Second, we would like the energy of the labeling \mathbf{x} defined above to be as close as possible to the capacity of the *st*-cut. This will allow us to obtain an accurate approximate solution \mathbf{x}' for problem (4) 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

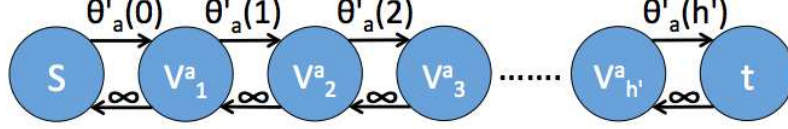


Figure 1. Arcs and their capacities for representing the unary potentials for the random variable X_a . According to the labeling defined in equation (5), if $x_a = \hat{x}_a$, then the arc (s, V_1^a) will be cut, which will contribute exactly $\theta_a(\hat{x}_a)$ to the capacity of the cut. If $x_a = s+i-1$ where $i \in \{1, \dots, h'-1\}$, 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 $(s, V_1^a) \cup \{(V_i^a, V_{i+1}^a), i = 1, \dots, h'-1\} \cup (V_{h'}^a, t)$ will be part of an st -cut with finite capacity, which will guarantee that we are able to obtain a valid labeling.

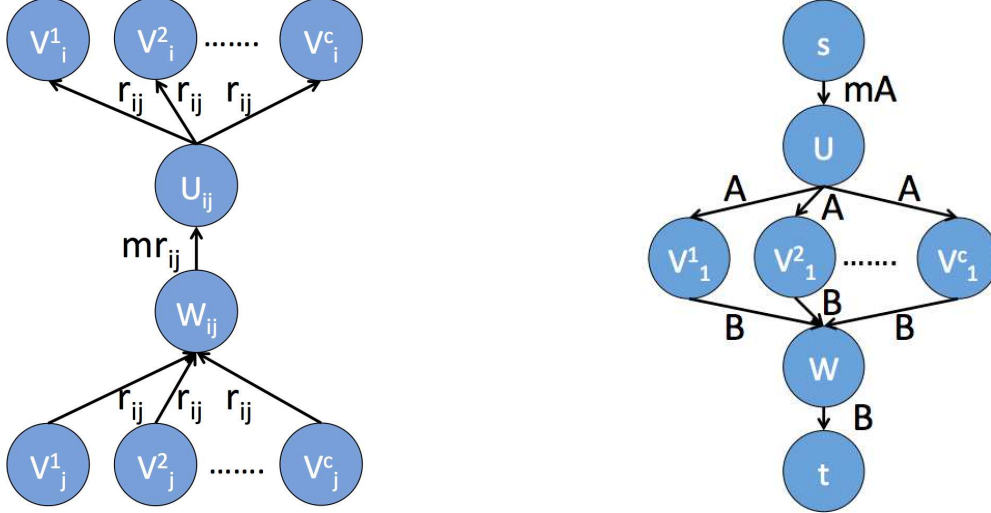


Figure 2. Arcs used to represent the high-order potentials for the clique $\mathbf{X}_c = \{X_1, X_2, \dots, X_c\}$. **Left.** The term r_{ij} is defined in equation (6). The arcs represent the sum of the m maximum convex distance functions over disjoint pairs of random variables when no random variable 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 (7). The arcs represent an overestimation of the clique potential for the case where some or all the random variables retain their old label (see supplementary material).

in Figure 1. Since all the unary potentials are non-negative, it follows that the arc capacities in Figure 1 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 $\mathbf{X}_c = \{X_1, X_2, \dots, X_c\}$. 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 2. For the arcs shown in Figure 2 (left), the capacities are specified using the term r_{ij} that is defined as follows:

$$r_{ij} = \begin{cases} \omega_c \frac{\bar{d}(i,j)}{2} & \text{if } i = j \\ \omega_c \bar{d}(i,j) & \text{otherwise.} \end{cases} \quad (6)$$

Here, the term $\bar{d}(i,j) = d(i-j+1) + d(i-j-1) - 2d(i-j) \geq 0$ since $d(\cdot)$ is convex, and $\omega_c \geq 0$ by definition. It follows that $r_{ij} \geq 0$ for all $i, j \in \{1, \dots, h'\}$. For the arcs shown in Figure 2 (right), the capacities are

specified using the terms A and B that are defined as follows:

$$A = M, B = \left(\omega_c M - \frac{\theta_c(\hat{\mathbf{x}}_c)}{m} \right). \quad (7)$$

Since $M \geq 0$ by definition, and $\theta_c(\hat{\mathbf{x}}_c) \leq \omega_c m M$ due to truncation, it follows that $A, B \geq 0$.

Given the above graph construction, an approximate solution \mathbf{x}' to problem (4) can be computed by first obtain the st -MINCUT, and then setting \mathbf{x}' according to equation (5). While it may not be immediately obvious why the above graph construction is suited to TCMC energy minimization, the following subsection provides a strong theoretical justification for its use by bounding the energy of the estimated labeling with respect to the optimal one.

5.2. Multiplicative Bounds

Similar to the case of pairwise TCM, we establish the accuracy of the range expansion algorithm by providing strong multiplicative bounds for special cases of

TMCM that are very useful in practice. The multiplicative bounds also serve to identify the best value of the interval length parameter h' .

Proposition 1. *The range expansion algorithm with $h' = M$ provides a multiplicative bound of $O(C)$ for the truncated max-of-linear model when $m = 1$. The term C is equal to the size of the largest clique. In other words, if \mathbf{x}^* is a labeling with the minimum energy and $\hat{\mathbf{x}}$ is the labeling estimated by the range expansion algorithm then*

$$\sum_{a \in \mathcal{V}} \theta_a(\hat{x}_a) + \sum_{c \in \mathcal{C}} \theta_c(\hat{\mathbf{x}}_c) \leq \sum_{a \in \mathcal{V}} \theta_a(x_a^*) + O(C) \sum_{c \in \mathcal{C}} \theta_c(\mathbf{x}_c^*). \quad (8)$$

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

Proposition 2. *The range expansion algorithm with $h' = \sqrt{M}$ provides a multiplicative bound of $O(C\sqrt{M})$ for the truncated max-of-quadratic model when $m = 1$.*

6. Discussion

We proposed a novel family of high-order random fields called truncated max-of-convex models (TMCM). The energy function of a TMCM consists of arbitrary unary potentials and high-order clique potentials that are proportional to the sum of the truncation of the m largest convex distance functions over disjoint pairs of random variables in the clique. In order to enable the use of TMCM, we developed a novel range expansion algorithm for energy minimization that retains the efficiency of *st*-MINCUT and provides provably accurate solutions.

From an applications point of view, our work opens up the possibility of improving the accuracy of other computer vision problems such as optical flow and segmentation using TMCM. From a theoretical point of view, our work can be thought of as a step towards the identification of graph representable submodular functions. We plan to investigate the properties of the Lovasz extension of submodular functions that enables the construction of an equivalent directed graph in an automated fashion.

References

- [1] C. Arora and S. Maheshwari. Multi label generic cuts: Optimal inference 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](#), [2](#), [3](#)
- [5] A. Delong, L. Gorelick, O. Veksler, and Y. Boykov. Minimizing energies with hierarchical costs. *IJCV*, 2012. [2](#)
- [6] A. Delong, A. Osokin, H. Isack, and Y. Boykov. Fast approximate energy minimization with label costs. In *CVPR*, 2010. [2](#)
- [7] P. Dokania and M. P. Kumar. Parsimonious labeling. In *ICCV*, 2015. [2](#)
- [8] A. Fix, C. Wang, and R. Zabih. A primal-dual algorithm for higher-order multilabel Markov random fields. In *CVPR*, 2014. [2](#)
- [9] B. Flach and D. Schlesinger. Transforming an arbitrary minsum problem into a binary one. Technical report, TU Dresden, 2006.
- [10] S. Gould, F. Amat, and D. Koller. Alphabet soup: A framework for approximate energy minimization. In *CVPR*, 2009. [2](#)
- [11] A. Gupta and E. Tardos. A constant factor approximation algorithm for a class of classification problems. In *STOC*, 2000. [1](#), [2](#), [3](#)
- [12] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *PAMI*, 2003.
- [13] P. Kohli, M. P. Kumar, and P. Torr. P3 & beyond: Solving energies with higher order cliques. In *CVPR*, 2007. [2](#)
- [14] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *CVPR*, 2008. [2](#)
- [15] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 2006. [1](#), [2](#), [3](#)
- [16] V. Kolmogorov. Minimizing a sum of submodular functions. *Discrete Applied Mathematics*, 2012. [2](#)
- [17] N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *ICCV*, 2007. [1](#), [2](#), [3](#)
- [18] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic MRFs. In *CVPR*, 2007. [1](#), [2](#), [3](#)
- [19] M. P. Kumar and P. Torr. Improved moves for truncated convex models. In *NIPS*, 2008. [1](#), [2](#), [3](#)
- [20] L. Ladicky, C. Russell, P. Kohli, and P. Torr. Graph cut based inference with co-occurrence statistics. In *ECCV*, 2010. [2](#)
- [21] R. Manokaran, J. Naor, P. Raghavendra, and R. Schwartz. SDP gaps and UGC hardness for multiway cut, 0-extension and metric labeling. In *STOC*, 2008. [1](#)
- [22] 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

[23] O. Veksler. Graph cut based optimization for MRFs with truncated convex priors. In *CVPR*, 2007. 1, 2, 3

7. Appendix - Proof of Proposition 1

Let f_m denote the labeling after the m -th iteration and $E(f_m)$ denote the corresponding energy. Also, f^* denotes optimal labeling of the MRF.

Let $r \in [0, L-1]$ be a uniformly distributed random integer and L be the length of the interval. Using r we define the following set of intervals

$$\Gamma_r = [0, r], [r+1, r+L], [r+L+1, r+2L], \dots, [., h-1]$$

where h is the total number of intervals.

Let $\mathbf{X}(f^*, I_m)$ contain all the random variables that take an optimal labeling in I_m , A_m be the set of all cliques for which all variables take optimum label in the interval I_m and B_m be the set of all cliques for which at least one, but not all, variable takes optimum label in the interval I_m .

The following equation can be deduced from the above definitions:

$$\sum_{x_a \in X} \theta_a(f^*(a)) = \sum \sum \theta_a(f^*(a)) \quad (9)$$

since $f^*(a)$ belongs to exactly one interval in I_r for all x . In order to make the analysis more readable, the following shorthand notation is introduced:

- We denote $\omega_c \max_{a,b \in X_c} d(f^*(a) - f^*(b))$ as $t_{ab,c}^m$
- We denote $\omega_c \max_{a \in X_c} d(f^*(a) - (i_m + 1))$ as $t_{a,c}^m$

At an iteration of our algorithm, given the current labeling f_m and an interval $I_m = [i_m + 1, j_m]$, the new labeling f_{m+1} obtained by solving the st-mincut problem reduces the energy by at least the following:

$$\begin{aligned} & \sum_{X_a \in \mathbf{X}(f^*, I_m)} \theta_a(f_m(a)) + \sum_{\mathbf{X}_c \in A_m \cup B_m} \theta_c(\mathbf{x}_c) \\ & - \left(\sum_{X_a \in \mathbf{X}(f^*, I_m)} \theta_a(f^*(a)) + \sum_{\mathbf{X}_c \in A_m} t_{ab,c}^m + \sum_{\mathbf{X}_c \in B_m} t_{a,c}^m \right) \end{aligned}$$

For the final label f of the range expansion algorithm, the above term should be non-positive for all intervals I_m because f is a local optimum. Hence,

$$\begin{aligned} & \sum_{X_a \in \mathbf{X}(f^*, I_m)} \theta_a(f(a)) + \sum_{\mathbf{X}_c \in A_m \cup B_m} \theta_c(\mathbf{x}_c) \\ & \leq \left(\sum_{X_a \in \mathbf{X}(f^*, I_m)} \theta_a(f^*(a)) + \sum_{\mathbf{X}_c \in A_m} t_{ab,c}^m + \sum_{\mathbf{X}_c \in B_m} t_{a,c}^m \right), \forall I_m \end{aligned}$$

We sum the above inequality over all $I_m \in \Gamma_r$. The summation of the LHS is at least $E(f)$. Also, using (9), the summation of the above inequality can be written as:

$$E(f) \leq \sum_{X_a \in \mathbf{X}} \theta_a(f^*(a)) + \sum_{I_m \in \Gamma_r} \left(\sum_{\mathbf{X}_c \in A_m} t_{ab,c}^m + \sum_{\mathbf{X}_c \in B_m} t_{a,c}^m \right)$$

We now take the expectation of the above inequality over the uniformly distributed random integer $r \in [0, L-1]$. The LHS of the inequality and the first term on the RHS (that is, $\sum \theta_a(f^*(a))$) are constants with respect to r . Hence, we get

$$E(f) \leq \sum_{X_a \in \mathbf{X}} \theta_a(f^*(a)) + \frac{1}{L} \sum_r \sum_{I_m \in \Gamma_r} \left(\sum_{\mathbf{X}_c \in A_m} t_{ab,c}^m + \sum_{\mathbf{X}_c \in B_m} t_{a,c}^m \right) \quad (10)$$

We have the following corollaries:

Corollary 1. When $d(\cdot)$ is linear, that is, $d(x) = |x|$, the following inequality holds true:

$$\begin{aligned} & \frac{1}{L} \sum_r \sum_{I_m \in \Gamma_r} \left(\sum_{\mathbf{X}_c \in A_m} t_{ab,c}^m + \sum_{\mathbf{X}_c \in B_m} t_{a,c}^m \right) \\ & \leq \max \left\{ \frac{c}{2} \left(2 + \frac{L}{M} \right), \left(2 + \frac{2M}{L} \right) \right\} \sum_{\mathbf{c} \in C} \theta_c(\mathbf{x}_c) \end{aligned} \quad (11)$$

Corollary 2. For the truncated max-of-linear metric, our algorithm obtains a multiplicative bound of $\frac{(c+2)+\sqrt{c^2+4}}{2}$ using $L = \frac{2-c+\sqrt{c^2+4}}{c}M$, where c is the size of the maximal clique. For $c = 2$, this gives a bound of $2 + \sqrt{2}$ using $L = \sqrt{2}M$.