

# Global Minimization for Continuous Multiphase Partitioning Problems Using a Dual Approach\*

Egil Bae<sup>†</sup>    Jing Yuan<sup>‡</sup>    Xue-Cheng Tai<sup>§</sup>

April 8, 2010

## Abstract

This paper is devoted to the optimization problem of continuous multi-partitioning, or multi-labeling, which is based on a convex relaxation of the continuous Potts model. In contrast to previous efforts, which are trying to tackle the optimal labeling problem in a direct manner, we first propose a novel dual model and then build up a corresponding duality-based approach. By analyzing the dual formulation, sufficient conditions are derived which shows that the relaxation is often exact, i.e. there exists optimal solutions that are also globally optimal to the original nonconvex Potts model. In order to deal with the nonsmooth dual problem, we suggest a smoothing method based on the log-sum exponential function and also indicate that such smoothing approach gives rise to the novel smoothed primal-dual model and suggests labelings with maximum entropy. Such smoothing method for the dual model produces a highly efficient expectation maximization algorithm for the multi-labeling problem, and provides a new thresholding scheme to obtain approximate solutions. Numerical experiments shows competitive performance in terms of quality and efficiency compared to several state of the art methods for the Potts model.

## 1 Introduction

The multi-partitioning problem, or multi-labeling problem, is the problem of assigning the unknown variable  $l$  pointwise at the image domain by a value chosen from the finite set  $\{l_1, \dots, l_n\}$  according to some model. Such kind of problems

---

\*Support from the Norwegian Research Council (eVita project 166075), National Science Foundation of Singapore (NRF2007IDM-IDM002-010) and Ministry of Education of Singapore (Moe Tier 2 T207B2202) are gratefully acknowledged.

<sup>†</sup>Department of Mathematics, University of Bergen, Norway ([Egil.Bae@math.uib.no](mailto:Egil.Bae@math.uib.no)).

<sup>‡</sup>Department of Computer Science, University of Western Ontario, Canada ([cn.yuanjing@googlemail.com](mailto:cn.yuanjing@googlemail.com))

<sup>§</sup>Department of Mathematics, University of Bergen, Norway and Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. ([tai@math.uib.no](mailto:tai@math.uib.no)).

appear extensively in the areas of image processing and computer vision. It is often formulated as the minimization of an energy function which mathematically encodes all the information needed for the imaging and vision task. The posteriori estimation is regarded as better as the energy is lower. In this work we focus on the Potts model, which is a special case of the labeling problem where the energy function does not favor any particular ordering between the labels.

One can solve such labeling problems in a manner of 'discrete' or 'continuous', depending on the spatial definition of  $l$ . When the labeling is given on a discrete grid, it reduces to the corresponding graphical model and its energy function is defined in terms of the 'cost' of corresponding nodes and edges of the graph, by the theory of Markov Random Fields. The node cost often evaluates the fidelity of given data. The edge cost, often pairwise [6, 23] or with high-order clique [20], measures the regularities of label assignment. In case of two labels such energy functions can be efficiently and globally minimized by graph cuts [13], provided they are submodular [23]. For certain problems with multiple labels this is also possible [16, 2]. However, in most cases with more than two labels the problem in the discrete context is NP hard. Therefore it can probably not be solved globally in polynomial time. The Potts model is an important such NP-hard discrete model, which can be seen as the natural extension of the partitioning problem to more than two labels. Several algorithms for finding suboptimal solutions to the Potts model or other such models exists, most notable are the graph-cut based alpha expansion and alpha-beta swap [6]. Moreover, by relaxing the constraint of discrete valued, most labeling problems can also be solved by linear programming approximately [24, 43]. Despite the efficiencies of such discrete approaches, their computation results are often constrained and biased by the discrete grid, i.e. the metrication error. By considering more neighbourhood nodes, such visual effects can be largely reduced [5, 21]. However, this often results in a great memory burden. Extensive applications based on graph cuts can be found in image processing and computer vision, such as image segmentation [6, 41], 3D reconstruction [22] etc.

Parallel to this development, variational methods have been proposed for solving some of the same problems in the continuous setting. Here one tries to divide a continuous domain  $\Omega$  into two or several subregions  $\{\Omega_i\}_{i=1}^n$  by minimizing certain energy functionals. The level set method [32] is an elegant tool for simplifying these energy functionals, such that numerical calculation is possible [10, 42]. A variant of the level set method called the piecewise constant level set method (PCLSM) [28, 27] expresses the energy in terms of a labeling function  $\phi$  and was the first to cast the labeling problem in the spatially continuous setting. The authors used the term "piecewise constant level set function" for this labeling function.

Fairly general energy functionals, such as the continuous variant of the Potts model, can be rewritten and minimized numerically by the level set methods. Unfortunately, the main disadvantage of both these variants is their potential of getting stuck in a possibly inferior local minima. For PCLSM the feasible minimization domain is non-convex and for more than two phases also the energy

functional itself is non-convex (but still locally convex). One notable exception in case of two phases is the approach of Nikolova et. al. [31]. By relaxing the binary constraint of the labeling function, the minimization problem becomes convex. They prove that thresholding the solution of the relaxed problem at almost any threshold between 0 and 1 yields the globally optimal solution to the original problem.

Recently, similar convex formulations have also been proposed for the multiphase continuous Potts model, by relaxing the integrality constraint of the labeling function. There is no proof that integer valued solutions exist to these problems. However, integer valued approximating solutions can be obtained by simple thresholding. Such approximate solutions may be closer to the exact global minimum than the local minima in the level set formulation. In [33] such a formulation was made using a labeling function by constraining the dual variables to a convex set and relaxing the integrality constraints of the primal variables. A primal-dual algorithm was developed. In [45, 25] the Potts model was instead formulated in terms of a binary vector function. A relaxation was proposed by ignoring the binary constraint and minimizing with respect to the vector function (primal variable) over the convex standard simplex. Some thresholding must be used to convert to an integer valued solution in a post processing step for both methods. Since no proof of optimality of the thresholded solution exists, the thresholded solution is simply accepted as suboptimal.

In the very recent work of Brown, Chan and Bresson [8] a method which was claimed to find global optimum for the continuous Potts model was proposed. Their idea is to apply convex functional lifting [34] to the piecewise constant level set method [27]. In short, this process introduces an extra dimension to make a non-convex energy functional convex. However, due to the extra dimension and the high order polynomials involved in PCLSM, we claim this approach is much more computationally expensive than our approach.

This paper builds on the work of Zach et. al. and Lellmann et. al. [45, 25]. The convex relaxed labeling problem is discussed under a primal-dual perspective. A novel dual model and algorithm is proposed by formulating the convex relaxed problem in terms of the dual variables only. This is in contrast to previous works which tackle the labeling functions directly. By analyzing the relaxed Potts model in terms of its dual formulation, sufficient conditions are derived which shows that optimal solutions are often expected to be binary; or equivalently, that optimal solutions are also globally optimal to the original non-convex Potts model. In order to deal with non-differentiability of the resulting energy functional of the dual model, a convex smoothing method is proposed, which also provides a new thresholding scheme to obtain approximate solutions. Other advantages of this approach are: it is based on a well-posed optimization problem in theory and leads to a reliable numerical scheme. The duality smoothing method avoids exploring the redundant simplex constraint of the primal labeling functions, e.g. [45, 25], and reduces the computational complexities greatly!

This paper is organized as follows: In Section 2 we introduce the continuous Potts model and its convex relaxation. We also present primal, primal-dual and

dual formulations of the convex relaxation of Potts model, which are equivalent to each other. Smoothed models are introduced in Section 3. Section 4 and 5 gives details on numerical implementation and numerical experiments are presented in Section 5.

## 2 Continuous Potts Model and its Convex Relaxation Approach

This work builds on the convex relaxation method of the nonconvex Potts model [45, 25]. We develop a novel dual formulation which allows us to give a deeper analysis of the relaxation. We show that optimal solutions of the relaxation is closely related to a global optimum of the original nonconvex Potts model. Furthermore, a novel duality-based algorithm is developed for the relaxed problem.

### 2.1 Continuous Potts Model

The continuous variant of the Potts model [35] partitions the continuous domain  $\Omega$  into  $n$  disjoint subdomains  $\{\Omega_i\}_{i=1}^n$  by minimizing

$$\begin{aligned} \min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} f_i(x) dx + \lambda \sum_{i=1}^n |\partial\Omega_i| \quad (1) \\ \text{s.t. } \cup_{i=1}^n \Omega_i = \Omega, \quad \Omega_k \cap \Omega_l = \emptyset, \forall k \neq l, \end{aligned}$$

where  $|\partial\Omega_i|$  measures the lengths of the boundaries of the disjoint subdomains  $\Omega_i$ ,  $i = 1, \dots, n$ . The functions  $f_i$ ,  $i = 1, \dots, n$ , defined on  $\Omega$  are given and evaluate the performance of label assignment at each partition  $\Omega_i$ . The Potts model was originally derived from statistical mechanics and formulated in the spatially discrete setting. Here, we have stated the corresponding definition in the continuous setting. Obviously, the Potts model favors the labelings with 'tight' or smooth boundaries. We will focus on image processing problems, in which  $\Omega$  is simply the image domain in 2D or 3D. In this case, the functions  $f_i$ ,  $i = 1, \dots, n$ , typically depend on the values of the input image  $I$ .

The Potts model has a close resemblance to the piecewise constant Mumford-Shah model [30]

$$\begin{aligned} \min_{\{\Omega_i\}_{i=1}^n, \mathbf{c}} \sum_{i=1}^n \int_{\Omega_i} |I - c_i|^p dx + \lambda \sum_{i=1}^n |\partial\Omega_i| \quad (2) \\ \text{s.t. } \cup_{i=1}^n \Omega_i = \Omega, \quad \Omega_k \cap \Omega_l = \emptyset, \forall k \neq l, \end{aligned}$$

where  $p = 1$  or  $2$  and  $I$  is the input image. In fact, when the constants  $\mathbf{c} = \{c_1, \dots, c_n\}$  are fixed, the Mumford-Shah model is a special case of the Potts model with  $f_i = |I - c_i|^p$ .

In order to compute the optimal partition, let  $u_i(x)$ ,  $i = 1, \dots, n$ , denote the characteristic functions of disjoint subdomains  $\Omega_i$  i.e.

$$u_i(x) = I_{\Omega_i}(x) := \begin{cases} 1, & x \in \Omega_i \\ 0, & x \notin \Omega_i \end{cases}, \quad i = 1, \dots, n.$$

The boundary lengths of the disjoint subdomains are given by

$$|\partial\Omega_i| = \int_{\Omega} |\nabla u_i| dx, \quad i = 1, \dots, n \quad (3)$$

The Potts model (1) can then be rewritten as

$$\min_{u_i(x) \in \{0,1\}} \sum_{i=1}^n \int_{\Omega} u_i(x) f_i(x) dx + \lambda \sum_{i=1}^n \int_{\Omega} |\nabla u_i| dx, \quad \text{s.t.} \quad \sum_{i=1}^n u_i(x) = 1. \quad (4)$$

Obviously, the Potts model (4) is non-convex due to the binary configuration of each function  $u_i(x)$ ,  $\forall x \in \Omega$ .

## 2.2 Convex Relaxation and Equivalent Models

We show the nonconvex Potts model (4) can be relaxed as a convex minimization problem which we call the primal model. We then give equivalent representations of the relaxation as a primal-dual model and a dual model.

### 2.2.1 Primal Model

The binary constraints of (4) were simply relaxed in [45], by minimizing  $u_i(x)$  over the interval  $[0, 1]$  instead. This gives rise to the convex relaxed Potts model defined as

$$\min_{(u_1(x), \dots, u_n(x)) \in \Delta_+} E^P(u) = \sum_{i=1}^n \int_{\Omega} u_i(x) f_i(x) dx + \lambda \sum_{i=1}^n \int_{\Omega} |\nabla u_i| dx \quad (5)$$

where the simplex constraint  $\Delta_+$  means for  $\forall x \in \Omega$ :

$$\Delta_+ = \{(u_1(x), \dots, u_n(x)) \mid \sum_{i=1}^n u_i(x) = 1; \quad u_i(x) \geq 0, \quad i = 1, \dots, n\}.$$

In Lellmann et al [25], the same relaxation was used, but with a slightly different regularization term  $\int \sqrt{|\nabla u_1(x)|^2 + \dots + |\nabla u_n(x)|^2} dx$ .

In this paper, we call the continuous optimization problem (5) *primal formulation* or *primal model* and  $u_i$ ,  $i = 1, \dots, n$ , *primal variables*, in comparison to its equivalent models discussed in later sections.

In addition to region based segmentation models like the Mumford-Shah model, edge based models like the geodesic active contour model [17] are powerful for many image processing problems. It was shown in [7, 39] that combining

regional and edge information of the image has many advantages. However, they only studied two-phase problems. Using the formulation outlined above, the combined model can be written with multiple phases

$$\min_{u_i \in \{0,1\}} \sum_{i=1}^n \int_{\Omega} \{u_i(x)f_i(x) + \lambda\tilde{g}(x)|\nabla u_i|\} dx, \text{ s.t. } \sum_{i=1}^n u_i(x) = 1. \quad (6)$$

Likewise, its convex relaxed version is

$$\min_{(u_1(x), \dots, u_n(x)) \in \Delta_+} \sum_{i=1}^n \int_{\Omega} \{u_i(x)f_i(x) + \lambda\tilde{g}(x)|\nabla u_i|\} dx. \quad (7)$$

The function  $\tilde{g}(x)$  is often called an edge indicator for problems like image segmentation. It takes small values at locations of large gradients or edges in the image. For a given image  $I$ , one possible choice is to take

$$\tilde{g}(x) = \frac{1}{c|\nabla I_{\sigma}(x)|^2 + 1},$$

where  $I_{\sigma}$  is a smoothed version of the input image  $I$ . There are also other alternatives for  $\tilde{g}(x)$ . If  $\tilde{g}(x)$  is chosen as the identity function, (6) reduces to the Potts model (4), and (7) corresponds to (5).

In this work, we discuss the convex relaxed Potts model (5) and its generalized model (7) as the approach to continuous multi-labelings and focus on (5), without loss of generalities.

## 2.2.2 Primal-Dual Model

By using integration by parts, it is well known that the total variation term in (5) can equivalently be formulated as a maximization problem

$$\lambda \int_{\Omega} |\nabla u| dx = \max_{p \in C_{\lambda}} - \int_{\Omega} \nabla u \cdot p dx = \max_{p \in C_{\lambda}} \int_{\Omega} u \operatorname{div} p dx \quad (8)$$

in terms of the dual variable  $p$  over the convex set  $C_{\lambda}$  defined as

$$C_{\lambda} := \{p : \Omega \mapsto \mathbb{R}^2 \mid \|p\|_{\infty} \leq \lambda, p_n|_{\partial\Omega} = 0\}, \quad (9)$$

see e.g. [29].

For the more general model (7),  $C_{\lambda}$  is given by

$$C_{\lambda} := \{p : \Omega \mapsto \mathbb{R}^2 \mid |p(x)|_2 \leq \tilde{g}(x), p_n|_{\partial\Omega} = 0\}. \quad (10)$$

By inserting such expression, the primal problem (5) can be identically formulated as

$$\min_{(u_1(x), \dots, u_n(x)) \in \Delta_+} \max_{p_i \in C_{\lambda}} E(u, p) = \sum_{i=1}^n \int_{\Omega} u_i(x)(f_i(x) + \operatorname{div} p_i(x)) dx. \quad (11)$$

The variables  $p_i$   $i = 1, \dots, n$  are named as *dual variables* in this paper. Therefore, the min-max problem (11) is called the equivalent *primal-dual formulation* or *primal-dual model* of (5), which can be optimized over both the primal variables  $u_i$  and the dual variables  $p_i$ .

Note that the min and max operators in the above primal-dual model (11) can be interchanged

$$\min_{(u_1(x), \dots, u_n(x)) \in \Delta_+} \max_{p_i \in C_\lambda} E(u, p) = \max_{p_i \in C_\lambda} \min_{(u_1(x), \dots, u_n(x)) \in \Delta_+} E(u, p) \quad (12)$$

because the conditions of the minimax theorem (see e.g., [11] Chapter 6, Proposition 2.4, also [12]) are all satisfied. That is,  $C_\lambda$  and  $\Delta_+$  are convex, and the energy function  $E(u, p)$  is linear to both variables  $u$  and  $p$ , hence convex l.s.c. for fixed  $p$  and concave u.s.c. for fixed  $u$ . This also implies the existence of at least one saddle point, see [11].

### 2.2.3 Dual Model

We will now derive another equivalent formulation of (5) by optimizing the primal-dual model (11) via the primal variable  $(u_1(x), \dots, u_n(x)) \in \Delta_+$  at each position  $x \in \Omega$ .

For the vector  $q = (q_1, \dots, q_n) \in \mathbb{R}^n$  and  $v = (v_1, \dots, v_n) \in \Delta_+$ , it is easy to verify that

$$\min_{(v_1, \dots, v_n) \in \Delta_+} \sum_{i=1}^n v_i q_i = \min(q_1, \dots, q_n), \quad (13)$$

Therefore, minimizing (11) over the primal variables  $u_i(x)$   $i = 1, \dots, n$ , at each position  $x \in \Omega$ , gives rise to

$$\max_{p_i \in C_\lambda} E^D(p) := \int_{\Omega} \{ \min(f_1 + \operatorname{div} p_1, \dots, f_n + \operatorname{div} p_n) \} dx. \quad (14)$$

We call (14) the *dual model* of (7).  $E^D(p)$  is called the dual energy functional.

By considering  $d(l_i, x) = f_i(x) + \operatorname{div} p_i(x)$ ,  $x \in \Omega$ , as the proximity measure of labeling  $x$  as  $l_i$ ,  $i = 1, \dots, n$ , the minimal distance indicates which label should be assigned at  $x$  by the dual model (14). In this sense, the dual formulation (14) can be viewed as a generalized center-based clustering formulation [40, 3], where  $f_i(x)$  are the data and  $\operatorname{div} p_i(x)$  are the centroids. In contrast to the classical clustering problem, the spatial centroids are formally constrained by some convex set. Moreover, updating  $\operatorname{div} p_i(x)$  associated with (8) minimizes the perimeter of the spatial partition  $\Omega_i$ ,  $i = 1, \dots, n$ , implicitly! This gives a geometrical explanation of the dual model (14) in the sense of minimal length clusterings.

### 2.2.4 Discussions and Comments

Now we have two equivalent optimization models, the primal-dual model (11) and the dual model (14), to the primal optimization problem (5). Clearly, the

energy functional  $E^P(u)$  of the primal model (5) is given by maximizing  $E(u, p)$  of the primal-dual problem (11) over the dual variable  $p$  first, i.e.

$$E^P(u) := \max_{p_i \in C_\lambda} E(u, p) = E(u, p^*),$$

where  $*$  indicates optimal variables.

Likewise, the energy functional  $E^D(p)$  of the dual model (14) is resulted by minimizing  $E(u, p)$  first over  $u$ , i.e.

$$E^D(p) := \min_{(u_1(x), \dots, u_n(x)) \in \Delta_+} E(u, p) = E(u^*, p).$$

As a consequence, we always have

$$E^P(u) \geq E(u, p) \geq E^D(p). \quad (15)$$

Let  $(u^*, p^*)$  be an optimal primal-dual pair, then by (12) we have

$$E^P(u^*) = E(u^*, p^*) = E^D(p^*).$$

Moreover, the equivalences between these models implies that we can also solve the convex relaxed Potts problem (5) by optimizing its dual model (14). In fact, when the  $n$  values  $(f_1(x) + \text{div } p_1^*(x), \dots, f_n(x) + \text{div } p_n^*(x))$  at  $x \in \Omega$  have a unique minimum, an optimal primal variable  $u^*(x)$  of  $\min_{u(x) \in \Delta_+} E(u, p^*)$  at  $x$  can be uniquely recovered, in view of (13), by

$$u_k^*(x) = \begin{cases} 1 & \text{if } k = \arg \min_{i=1, \dots, n} (f_i(x) + \text{div } p_i^*(x)) \\ 0 & \text{otherwise} \end{cases}, \quad (16)$$

which is a binary indicator vector and exactly indicates the optimal label function  $u^*$  at position  $x$  to be nonzero only at the  $k$ th component. Clearly, such binary  $u^*(x)$  is globally optimal both to the convex relaxed Potts model (5) and the nonconvex Potts model (4).

Based on the above consideration, we propose such a duality-based approach by maximizing the dual functional  $E^D(p)$ , in contrast to previous works which tackle the primal unknowns  $u_i$ ,  $i = 1, \dots, n$ , directly. Moreover, the dual model (14) also provides a powerful tool to analyze the connections between the global optimums of the non-convex Potts model (4) and its relaxed version (5).

### 2.3 Global Optimums of Convex Relaxed Potts Model

The existence of a global binary optimum of the convex relaxed formulation (5), i.e. the exactness of (5), is still open. However, we can show (5) is exact under specified conditions. To do this, we first state the consistency between the minimum  $u^*$  of (5) and the maximum  $p^*$  of its dual model (14) as follows

**Theorem 1.** Given a maximum  $p^*$  of the dual problem (14), and let  $u^*$  be an optimum to the primal problem (5) such that  $(u^*, p^*)$  is an optimal primal-dual pair (11). If the values of  $(f_1(x) + \text{div } p_1^*(x), \dots, f_n(x) + \text{div } p_n^*(x))$ , at some

$x \in \Omega$ , have a unique minimum, e.g.  $f_k(x) + \operatorname{div} p_k^*(x)$ , then the value of  $u^*(x)$  at  $x$  must be

$$u_k^*(x) = 1 \quad \text{and} \quad u_i^*(x) = 0, \quad i \neq k. \quad (17)$$

If the values of  $(f_1(x) + \operatorname{div} p_1^*(x), \dots, f_n(x) + \operatorname{div} p_n^*(x))$  at some  $x \in \Omega$  have  $k > 1$  minimums, e.g.  $f_j(x) + \operatorname{div} p_j^*(x)$   $j \in \{t_1, \dots, t_k\}$ , then the value of  $u^*(x)$  at  $x$  must satisfy

$$\sum_{i=1}^k u_{t_i}^*(x) = 1 \quad \text{and} \quad u_j^*(x) = 0, \quad j \notin \{t_1, \dots, t_k\}. \quad (18)$$

*Proof.* Let  $u^d = (u_1^d, \dots, u_n^d)$  be any vector consistent with  $p^*$ , in terms of (17) or (18) for  $\forall x \in \Omega$ . Clearly

$$E^D(p^*) = E(u^d, p^*) = E^P(u^d) \quad (19)$$

as  $p^*$  is an optimum of the dual problem (14).

If a global optimum  $u^*$  of (5) at  $x \in \Omega$  is not consistent with  $p^*$  in the sense of (17) or (18), then we must have

$$\sum_{i=1}^n u_i^d(x)(f_i(x) + \operatorname{div} p_i^*(x)) < \sum_{i=1}^n u_i^*(x)(f_i(x) + \operatorname{div} p_i^*(x)),$$

and

$$E(u^d, p^*) < E(u^*, p^*).$$

Therefore, by the fact (15),

$$E^D(p^*) = E(u^d, p^*) < E(u^*, p^*) \leq E^P(u^*).$$

In this regard, (19) indicates

$$E^P(u^d) < E^P(u^*).$$

It contradicts the fact that  $u^*$  is a global optimum of the primal problem (5).  $\square$

Then it follows directly, from Theorem 1, that

**Proposition 2.** Let  $p^*$  be one optimum of the dual problem (14), if the values  $(f_1(x) + \operatorname{div} p^*(x), \dots, f_n(x) + \operatorname{div} p^*(x))$  have a unique minimum at all  $x$  in  $\Omega$ , then any global minimum  $u^*$  of the primal problem (7), where  $(u^*, p^*)$  is an optimal primal-dual pair, must be binary. Therefore,  $u^*$  is also a global optimum to the original non-convex Potts model (4).

*Proof.* It is obvious to verify the first part in view of (17).

The last part follows from the fact that the feasible set of the convex relaxed Potts model (5) contains the feasible set of the nonconvex Potts model (4). Therefore, any global binary optimum of the convex relaxed Potts model, which is feasible in the nonconvex Potts model (5), is also globally optimal to the nonconvex Potts model.  $\square$

In addition, we have the following corollary

**Corollary 3.** Let  $p^*$  be one optimum of the dual problem (14), if the values of  $(f_1(x) + \operatorname{div} p_1^*(x), \dots, f_n(x) + \operatorname{div} p_n^*(x))$  have a unique minimum at almost every  $x \in \Omega$ , then any global minimum  $u^*$ , where  $(u^*, p^*)$  is an optimal primal-dual pair of the primal problem (5) or (7) is unique and binary almost everywhere in  $\Omega$ .

In the case where there are two minimums of  $(f_1(x) + \operatorname{div} p_1^*(x), \dots, f_n(x) + \operatorname{div} p_n^*(x))$  at some positions  $x \in \Omega$ , the existence of a binary solution can also be given by

**Proposition 4.** Given a closed subdomain  $\tilde{\Omega} \subset \Omega$  and an optimum  $p^*$  of the dual problem (14), if for any  $x \in \tilde{\Omega}$ , the values of  $(f_1(x) + \operatorname{div} p_1^*(x), \dots, f_n(x) + \operatorname{div} p_n^*(x))$  have two minimums (not unique), say  $f_k(x) + \operatorname{div} p_k^*(x)$  and  $f_j(x) + \operatorname{div} p_j^*(x)$ , and for any  $x \in \Omega \setminus \tilde{\Omega}$ , the values of  $(f_1(x) + \operatorname{div} p_1^*(x), \dots, f_n(x) + \operatorname{div} p_n^*(x))$  have only one unique minimum, then there exist global optimums  $u^*$  of the primal problem (5), which are binary.

The proof of Prop. 4 is given in the appendix. The above proposition can be easily extended to the case when there are multiple such subdomains and for each subdomain there are two minimums. As a corollary, this shows there always exists a global binary minimizer for two phase problems where  $n = 2$ . This corollary is linked to the result of Chan et. al. [31] and Strang et. al. [38], who gave a different proof of existence of binary minimizers for relaxed two phase problems.

For the general case, where the number of minimums is greater than two at some positions  $x \in \Omega$  a strict proof of existence of binary solutions will be considered in a future work.

Optimal solutions  $p^*$  to the dual model (14) are generally not unique. Similarly, optimal primal solutions  $u^*$  to (5) are generally not unique, there may exist both binary and non-binary solutions  $u^*$  that are optimal. In order to deal with non-differentiability of the dual energy functional (14), a smooth approximation functional is introduced in the next section. For this smooth model we show there exists a unique optimal dual solution. For reasons we cannot fully explain, this smooth model always favors solutions  $p^*$  where at each  $x \in \Omega$ ,  $(f_1(x) + \operatorname{div} p_1^*(x), \dots, f_n(x) + \operatorname{div} p_n^*(x))$  has a unique minimum. Therefore binary primal solutions can be recovered by the scheme (16). Since there is an approximation between the smooth and non-smooth models, one cannot always guarantee the obtained solutions are exact.

### 3 Smoothed Models

In order to solve the optimal labeling problem (5) through its nonsmooth dual model (14), we propose a smoothing method in this section, which leads to the *smoothed primal-dual model* and *smoothed dual model*, associated with (11) and (14). Especially, the smoothed dual model also gives rise to a simple and efficient numerical algorithm which is proposed in the following section.

### 3.1 Asymptotic Function and Smoothed Dual Model

We first introduce the asymptotic function in order to derive the smoothing method. The asymptotic function  $g_\infty$  of a proper convex function  $g(u)$  is also a proper convex function, positively homogeneous and defined in an approximation way [36, 40] as

$$g_\infty(z) = \lim_{s \rightarrow 0^+} \{g_s(z) := sg(s^{-1}z)\}.$$

For example,

$$g(u) = \sqrt{1 + \|u\|^2}, \quad g_\infty(z) = \|z\|;$$

and

$$g(u) = \log \sum_{j=1}^k e^{u_j}, \quad g_\infty(z) = \max_{1 \leq j \leq k} z_j. \quad (20)$$

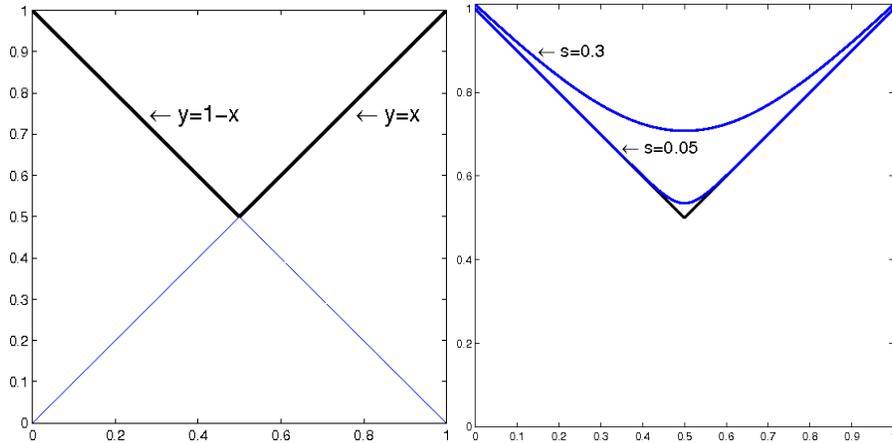


Figure 1: The left graph, the black and bold line, gives the function  $f(x) = \max(1 - x, x)$ , at the section  $x \in [0, 1]$ . The right graph shows the approximation of  $f(x)$  by the Log-Sum exponential function  $f_s(x) = s \log(\exp((1 - x)/s) + \exp(x/s))$  where  $s = 0.3$ : the upper blue line,  $s = 0.05$ : the lower blue line.

We use an example to show the smoothing effects of the Log-Sum exponential function (20) for the highly nonsmooth function  $\max_{1 \leq j \leq k} z_j$ . In Fig. 1, the nonsmooth function  $f(x) = \max(1 - x, x)$  is given in the first graph. We use the Log-Sum exponential function (20) to approximate it by

$$f_s(x) = s \log(\exp((1 - x)/s) + \exp(x/s)),$$

where  $s > 0$ . We see, by the two blue lines on the right graph, that the approximation becomes better when  $s$  is chosen smaller.

Likewise, we apply (20) to approximate the min function in (14) by choosing a small parameter  $s > 0$ . In this way, the nonsmooth optimization problem (14) can be approximated by

$$\max_{p_i \in C_\lambda} E_{s>0}^D(p) := -s \int_{\Omega} \left\{ \log \sum_{i=1}^n \exp\left(\frac{-f_i - \operatorname{div} p_i}{s}\right) \right\} dx. \quad (21)$$

We call the new optimization problem (21) the *smoothed dual model* in comparison to the original dual one (14).

In contrast to the non-smooth dual model (14), the optimum of the smoothed dual model (21) is always unique due to the strict concaveness of the energy function of (21).

### 3.2 Equivalent Smoothed Models and Maximum Entropy Labelings

Actually, it is well known that the smooth log-sum function has an identical expression [36]:

**Lemma 5.** For any given  $\mu \in \Delta_+$  and  $h \in \mathbb{R}^n$ ,

$$\log \sum_{i=1}^n \mu_i e^{h_i} = \max_{u \in \Delta_+} \left\{ \langle u, h \rangle - \sum_{i=1}^n u_i \log \frac{u_i}{\mu_i} \right\}.$$

Let  $\mu_i = 1/n, i = 1 \dots n$ . By the results of lemma 5, we see that the *smoothed dual model* (21) is just equivalent to

$$\max_{p_i \in C_\lambda} \min_{u(x) \in \Delta_+} E_s(u, p) = \int_{\Omega} \left\{ \sum_{i=1}^n u_i (f_i + \operatorname{div} p_i) + s \sum_{i=1}^n u_i \log u_i \right\} dx. \quad (22)$$

In view of the primal-dual model (11), the energy functional in the optimization problem (22) is just the energy functional of (11) plus an entropy-penalizing term. Such entropy penalization provides a proper regularization or smoothing of the original function. We, likewise, call the optimization problem (22) the *smoothed primal-dual model*.

Correspondingly, optimizing the dual variables  $p$  in (22) leads to the equivalent *smoothed primal model*:

$$\min_{u(x) \in \Delta_+} E_s^P(u) = \int_{\Omega} \left\{ \sum_{i=1}^n u_i f_i + \lambda \sum_{i=1}^n |\nabla u_i| + s \sum_{i=1}^n u_i \log u_i \right\} dx. \quad (23)$$

Clearly, the positive value  $s$  here works as a penalization parameter. When  $s$  approaches 0, the optimization problem (22) approaches the original primal-dual problem (11) and the smoothed primal model (23) approaches the nonsmooth version (5). At this, the smoothed primal-dual model (22) shares the same formulation of the *maximum entropy clustering algorithms* [37]. To this end, we also call our smoothing approach given by (21) or (22) the method of *maximum entropy labelings*.

### 3.3 Approximation Bounds of Smoothed Models

In fact, the Log-Sum exponential function gives the following approximation bound of the maximum function  $\max_{1 \leq i \leq k} z_i$  [40]

**Lemma 6.** For each  $\mu \in \Delta_+$ , the following inequalities hold,

$$\sum_{i=1}^k \mu_i z_i \leq \log \sum_{i=1}^k \mu_i e^{z_i} \leq \max_{1 \leq i \leq k} z_i .$$

Moreover, for  $s > 0$

$$\sum_{i=1}^k \mu_i z_i \leq \lim_{s \rightarrow 0^+} \left\{ s \log \sum_{i=1}^k \mu_i e^{z_i/s} \right\} \leq \max_{1 \leq i \leq k} z_i .$$

Proof is referred to [40].

Then in view of Lemma 6, we have the approximation bound of the smoothed dual model:

**Proposition 7.** For any  $s > 0$ , the smoothed dual model (21) gives an approximation of (14), which has the bound:

$$0 \leq E^D(p) - E_{s>0}^D(p) \leq s \log n |\Omega|$$

where the functions  $E^D(p)$  and  $E_{s>0}^D(p)$  are the energy functional of (14) and (21) respectively,  $|\Omega|$  is the area of the domain  $\Omega$ .

*Proof.* Define the function  $G_s(x)$  as

$$G_s(x) := -s \log \left\{ \sum_{i=1}^n \exp\left(\frac{-f_i(x) - \operatorname{div} p_i(x)}{s}\right) \right\},$$

i.e. the component function of (21) to be integrated.

Let  $\mu_i = 1/n$  and  $z_i = -(f_i(x) + \operatorname{div} p_i(x))$  for each  $x \in \Omega$ . By Lemma 6, we have

$$\sum_{i=1}^n \frac{f_i(x) + \operatorname{div} p_i(x)}{n} \geq s \log n + G_s(x) \geq \min_{1 \leq i \leq k} (f_i(x) + \operatorname{div} p_i(x)).$$

Therefore,

$$\min_{1 \leq i \leq k} (f_i(x) + \operatorname{div} p_i(x)) - G_s(x) \leq s \log n ,$$

and

$$E^D(p) - E_{s>0}^D(p) \leq s \log n |\Omega| .$$

On the other hand, through lemma 5 and (22), we have

$$G_s(x) = \min_{u \in \Delta_+} \sum_{i=1}^n u_i (f_i + \operatorname{div} p_i) + s \sum_{i=1}^n u_i \log u_i .$$

Hence

$$0 \leq -s \sum_{i=1}^n u_i \log u_i \leq \min_{1 \leq i \leq k} (f_i(x) + \operatorname{div} p_i(x)) - G_s(x);$$

then

$$E^D(p) - E_{s>0}^D(p) \geq 0.$$

□

By Prop. 7, the approximation bound of the smoothed model (21) depends on the smoothing parameter  $s$ . Hence by choosing  $s$  small enough, the smoothed dual model (21) solves the original nonsmooth dual model (14) within an expected error bound.

## 4 Algorithm

The smooth energy function considered in the smoothed dual model (21) provides the feasibility to build up an efficient and simple numerical scheme over dual variables  $p_i(x)$ ,  $i = 1 \dots n$ . In order to maximize the energy functional of (21), we propose a projected gradient algorithm, see Alg. 1, which contains the same steps as the algorithms suggested in [9, 18].

---

### Algorithm 1 Projection-based Smoothing Algorithm

---

- Let  $\delta > 0$  be chosen as some suitable step-size and let  $p_i^0$ ,  $i = 1, \dots, n$  be chosen as the starting values, set  $k = 0$  then start;

- Compute

$$u_i^k = \frac{e^{\frac{-f_i - \operatorname{div} p_i^k}{s}}}{\sum_{i=1}^n e^{\frac{-f_i - \operatorname{div} p_i^k}{s}}}, \quad i = 1, \dots, n; \quad (24)$$

- Update  $p_i^{k+1}$ ,  $i = 1, \dots, n$  by

$$p_i^{k+1} = \operatorname{Proj}_{C_\lambda}(p_i^k + \delta \nabla u_i^k), \quad i = 1, \dots, n,$$

where  $\operatorname{Proj}_{C_\lambda}$  is the projection operator to the convex set  $C_\lambda$ ;

- Let  $k = k + 1$  and restart  $k + 1$  iteration until convergence. When convergence is achieved, the primal variable  $u$  is recovered by

$$u_l = \begin{cases} 1 & \text{if } l = \arg \min_{i=1, \dots, n} (f_i + \operatorname{div} p_i) \\ 0 & \text{otherwise.} \end{cases}.$$

---

The two main steps at each iteration can also be explained as the Expectation Maximization (EM) steps:

- **Expectation Step**, compute the conditional probabilities by fixing the dual variables  $p_i^k$ ,  $i = 1 \dots n$ :

$$u_i^k = \frac{e^{\frac{-f_i - \operatorname{div} p_i^k}{s}}}{\sum_{i=1}^n e^{\frac{-f_i - \operatorname{div} p_i^k}{s}}}, \quad i = 1, \dots, n;$$

- **Maximization Step**, maximize the energy functional by fixing  $u_i^k$ ,  $i = 1 \dots n$ :

$$p_i^{k+1} = \operatorname{Proj}_{C_\lambda}(p_i^k + \delta \nabla u_i^k), \quad i = 1, \dots, n.$$

The above *maximization step* is implemented by the following projected descent steps:

- **Gradient-Descent Step**, compute

$$\tilde{p}_i^{k+1} = p_i^k + \delta \nabla u_i^k, \quad i = 1, \dots, n$$

where  $\nabla u_i^k$  is the gradient of the energy functional of (21)

- **Projection Step**, compute the projection to the convex set  $C_\lambda$ :

$$p_i^{k+1} = \operatorname{Proj}_{C_\lambda}(\tilde{p}_i^{k+1}), \quad i = 1, \dots, n.$$

#### 4.1 Analysis on Smoothing Algorithm

Now we give analytical results on the proposed Alg. 1.

**Proposition 8.** Let  $p^s$  be optimal to the smoothed dual model (21) and  $u^s$  be given by  $p^s$

$$u_i^s = \frac{e^{\frac{-f_i - \operatorname{div} p_i^s}{s}}}{\sum_{i=1}^n e^{\frac{-f_i - \operatorname{div} p_i^s}{s}}}, \quad i = 1, \dots, n. \quad (25)$$

$(u^s, p^s)$  is the unique optimum of the smoothed primal-dual model (22), also  $u^s$  is the unique minimum of the smoothed primal model (23).

*Proof.* By inserting the expression (25) into the smoothed primal-dual model (22) we get

$$E_s(u^s, p^s) = -s \int_{\Omega} \left\{ \log \sum_{i=1}^n \exp\left(\frac{-f_i - \operatorname{div} p_i}{s}\right) \right\} dx = E_s^D(p^s).$$

Hence, this also implies that  $u^s$  solves the smoothed primal model (23) due to smoothness and equivalences of the primal and dual models.

For each  $x \in \Omega$ , we denote

$$h(u_1, \dots, u_n) = \sum_{i=1}^n u_i (f_i + \operatorname{div} p_i^s) + s \sum_{i=1}^n u_i \log u_i,$$

i.e.  $E_s(u^s, p^s) = \int_{\Omega} h(u_1^s(x), \dots, u_n^s(x)) dx$ . Then we have

$$\frac{\partial^2 h}{\partial u_i^2} = s \frac{1}{u_i} > 0.$$

Hence  $h$  is strictly convex which implies that for any  $u' \neq u^s \in \Delta_+$ ,  $E_s(u', p^s) > E_s(u^s, p^s) = E_s^D(p^s)$ . Therefore, since  $p_s$  is unique, it follows that  $(u^s, p^s)$  is the unique solution to the smoothed primal-dual model (22).  $\square$

As seen in Prop. 8, the expectation step (24) in Alg. 1 just gives an approximation of  $u^s$ . Obviously,  $u^s$  may not be binary. In order to recover a binary solution, Alg. 1 finally uses a threshold to estimate an optimum of the nonsmooth primal problem (5), i.e.

$$\tilde{u}_k^s(x) = \begin{cases} 1 & \text{if } k = \arg \min_{i=1, \dots, n} (f_i(x) + \text{div } p_i^s(x)) \\ 0 & \text{otherwise} \end{cases}. \quad (26)$$

Observe (25) and the nondecreasing exp. function, (26) just leads to

$$\tilde{u}_k^s(x) = \begin{cases} 1 & \text{if } k = \arg \max_{i=1, \dots, n} u_i^s(x) \\ 0 & \text{otherwise} \end{cases}. \quad (27)$$

We contribute the rest of this section to analyze such thresholding step.

Assume  $p^s$  is the optimum of the smoothed dual problem (21). Let  $d^s(x)$  be the gap between the second smallest and smallest component of  $\{f_i(x) + \text{div } p_i^s(x)\}_{i=1}^n$ , i.e.

$$d^s(x) = \min_i^{2\text{nd}} (f_i(x) + \text{div } p_i^s(x)) - \min_i (f_i(x) + \text{div } p_i^s(x)), \quad (28)$$

where  $\min^{2\text{nd}}$  denotes the second smallest value, clearly  $d^s(x) \geq 0$ . We denote  $D^s = \min_{x \in \Omega} d^s(x)$ .

We show that when the smoothing parameter  $s$  is smaller than some bound in case  $D^s > 0$ , the thresholded solution  $\tilde{u}^s$  computed by Alg. 1 gives the exact global optimum of the nonsmooth primal model (5), and therefore also of the original nonconvex Potts model (4) (see Prop. 2).

**Proposition 9.** Let  $p^s$  be the optimum of the smoothed dual model (21) and the smoothing parameter  $s > 0$  be chosen. In case  $D^s > 0$  and

$$s \leq \frac{D^s}{\log [\gamma(n-1)/(1-\gamma)]} \quad (29)$$

where  $\gamma \in [0.5, 1)$ , the thresholded result  $\tilde{u}^s$  given by Alg. 1 exactly solves the convex relaxed primal model (5), therefore the original nonconvex Potts model (4), in a global manner.

Proof is given in the Appendix B.

In fact,  $D^s > 0$  just indicates that there exists an unique minimum of  $f_i(x) + \text{div } p_i^s(x)$  for all  $x \in \Omega$  and the global optimizer of the nonconvex Potts model

can be given by its convex relaxed version, see Prop. 2. Therefore, the bound of  $s$  given in Prop. 9 presents a condition under which the optimum computed by the smoothed model also solves the nonsmooth model (5) and the nonconvex Potts model (4) globally and exactly!

When the relationship (29) between  $s$  and  $D^s$  is not satisfied, the algorithm still provides a binary approximate solution. Numerical experiments confirm that such suboptimal solutions are indeed very close to a global minimum.

## 5 Numerical Experiments

In this work, we apply the *mimetic finite-difference method* [15, 14] to build up relevant discretization scheme. A 2-D square grid where the size of each grid cell is assumed to be 1 is used. By this, 2-D scalar fields and vector fields are given by their discrete representations with the *mimetic finite-difference method* and four types of discrete 2-D fields are summarized on this 2-D grid to model various fields and mimic continuous vector calculus in discrete settings. The definitions of these four corresponding linear function spaces are listed below, see also Fig. 2:

- $H_V$ : the space of *scalar fields* defined on cells: the value of the scalar field is given at the center of each cell (see the empty circles of Fig. 2);
- $H_P$ : the space of *scalar fields* defined on vertices: the value of the scalar field is given at each vertex (see the filled circles of Fig. 2);
- $H_E$ : the space of *vector fields* defined tangential to sides: the value of the vector field is given at the center of each side of cells and parallel to the hosting side (see the related sides of Fig. 2);
- $H_S$ : the space of *vector fields* defined normal to sides: the value of the vector field is given at the center of each side of cells and normal to the hosting side (see the related sides of Fig. 2).

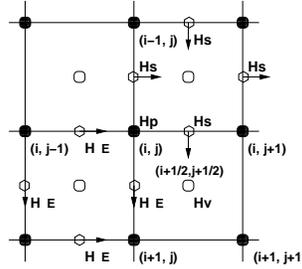


Figure 2: Definitions of finite-dimensional spaces of scalar fields,  $H_V$  and  $H_P$ , and vector fields,  $H_E$  and  $H_S$ , on a 2-D square grid.

We implement our numerical scheme mainly by applying the  $H_V$  space for 2-D scalar fields and the  $H_S$  space for vector fields. We refer to [44] for detailed implementation and cell-wise computation of functions.

We demonstrate the performance of the smoothed dual model by several experiments and compare with established methods. Alpha expansion and alpha-beta swap [6] are widely considered state of the art for approximately minimizing the discrete version of (1) with anisotropic total variation (TV) term. The method proposed in this paper instead minimizes the more ideal energy functional with isotropic TV term, i.e. the euclidian length of the boundaries. Because of this difference, energy comparison is not straight forward. However, there exists a result which allows to approximate the euclidian curve length on a discrete grid. This result is called the Cauchy-Crofton formula and was specialized for computer vision problems in [4]. In short, it gives a formula for edge weights between neighboring grid points such that the discrete boundary length converges to the euclidian boundary length as the mesh size goes to zero and the number of neighbors goes to infinity. This result can therefore be used to determine weights on regularization edges in the discrete model, such that it correctly corresponds to the continuous model. It is also used to compute the final energy of the outputs produced by the different methods. Secondly, we evaluate quality and efficiency with the approaches of [45, 25]. Energy plots for all experiments can be found in Figure 10. The final energies of the different methods are plotted as a function of the regularization parameter  $\lambda$ . Since the final energy is eventually evaluated on a discrete grid there will be some bias in favor of the graph cut based approaches. Some comparisons are also made to the very recent convex relaxation approach [34] for minimizing the isotropic variant of the energy functional, however an extensive experimental comparison with this approach is out of the scope of this paper.

In experiments where the correct solution is known, we have also compared the percentage of misclassified pixels, Table 1. The regularization parameter  $\lambda$  has here been manually selected for each method to minimize the percentage of misclassified pixels. . The implementation of the proposed method is made in matlab and the implementations of alpha expansion and alpha-beta swap are made in C++ [6]. The input images in Figure 3 and 8 was first used by Pock et. al. [33], and the input images in Figure 6 and 7 was first used by Lellmann et. al. [25].

## 5.1 Qualitative evaluation

In Figure 3, 10 classes have been used, with color data fidelity

$$f_i = \sum_{j=1}^3 |I - c_i^j|, \quad i = 1, \dots, 10, \quad (30)$$

where  $\{c_i\}_{i=1}^{10}$  are predefined color vectors. For this rather difficult problem with a large number of labels, the smoothed dual model clearly outperforms the graph cut based approaches in terms of energy, see Figure 10 (a). In Figure 3 (c) we



Figure 3: (a) Input, (b) alpha expansion 4 neighbors, (c) alpha expansion 8 neighbors, (d) Pock et. al. (e) dual model.

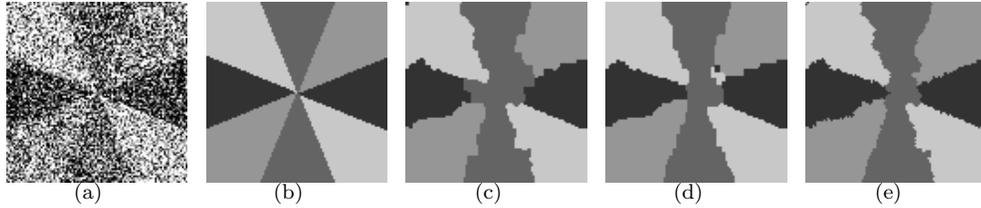


Figure 4: Experiment 1: (a) Input, (b) ground truth, (c) alpha expansion, (d) alpha-beta swap, (e) dual model. Size  $100 \times 100$ .

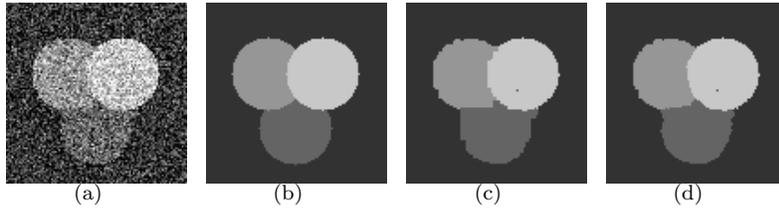


Figure 5: Experiment 2: (a) Input, (b) ground truth, (c) alpha expansion, (d) dual model. Size:  $100 \times 100$ .

compare with the recent method of Pock et. al. Their method seems to recover almost integer valued solutions up to some blurring of the boundaries. Our method, on the other hand, yields results that are integer valued everywhere.

Some artificial examples are presented next in experiment 1-4, Figure 4 - 7. The leftmost gray scale image  $I$  is to be classified into 4 classes by using the  $L^1$  norm in the data fidelity term

$$f_i = |I - c_i|, \quad i = 1, \dots, 4, \quad (31)$$

where  $\{c_i\}_{i=1}^4$  are predefined real values. We observe that in experiment 1,2 and 4 the new method with  $s = 0.01$  outperforms alpha expansion and alpha-beta swap, both in terms of visual quality and number of misclassified pixels (Table 1). In experiment 3, alpha expansion performs best. This is due to the fact that the correct solution only consists of horizontal boundaries which are favored by the anisotropic model. However, the proposed method outperforms alpha-beta swap and the primal model for this example. In experiment 4, where the boundaries are diagonal, the dual model clearly performs best. For energy plots, see Figure 10. In terms of energy, our approach performs about equally well as alpha expansion for these two examples. Observe also that our approach can obtain solutions of lower energy than approaches [45, 25]. This is particularly visible in Figure 10(c).

The advantage of the smoothing is illustrated in the next example, Figure

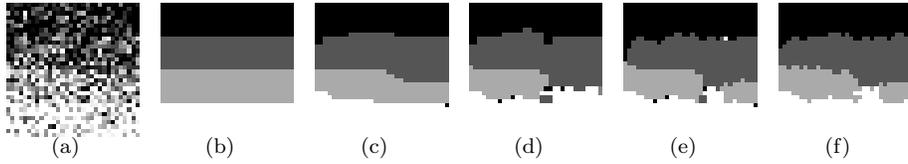


Figure 6: Experiment 3: (a) Input, (b) ground truth, (c) alpha expansion, (d) alpha-beta swap, (e) Lellmann et. al., (f) dual model. Size:  $32 \times 32$ .

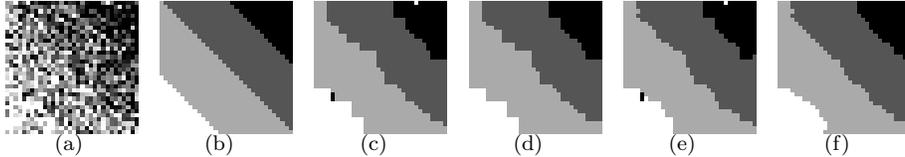


Figure 7: Experiment 4: (a) Input, (b) ground truth, (c) alpha expansion, (d) alpha-beta swap, (e) Lellmann et. al., (f) dual model. Size:  $32 \times 32$ .

8, where we want to recover a triple junction by filling in the gray area. The data term is given by  $f_i = 0$  for  $i = 1, 2, 3$  inside the gray disk, and by the color distance (30) outside the gray disk. This is a typically difficult example as the data term is equal for all labels. The global minimum of Potts model will fill in the gray area such that the total length of the boundaries between the labels are minimized, i.e. the boundaries meet with 120 degree angles in the center. For this example we expect that for the non-smooth model ( $f_1(x) + \text{div } p_1^*(x), \dots, f_n(x) + \text{div } p_n^*(x)$ ) does not have a unique minimum for some points inside the gray area, which makes it difficult to determine the label at such points. However, for the smooth model a unique minimum can be obtained at each point. The value of  $D_s$  converges to around  $10^{-9}$  for this example when  $s = 0.005$ . Therefore condition (29) is not satisfied, so we cannot guarantee an exact solution. In spite of this, the thresholding scheme still provides approximate binary solution which in fact coincides with the exact global minimum. The final result is shown in Figure (8), where we also compare with other methods. The approach of Lellmann et. al. does not recover a binary solution. Alpha expansion yields a binary, but incorrect result, Figure (8) (d). It can easily be seen geometrically that this is a local optimum, i.e. no alpha expansion move can yield a result of lower energy. We also compare with the convex relaxation of Pock et. al. [33], who first tested their method on this image. As can be seen, they are not able to recover the integer valued global minimum, although they are close for this particular example. Numerical calculations for triple junctions have also been tested in [26] showing that the piecewise constant level set method is able to produce 120 degrees for the junctions.

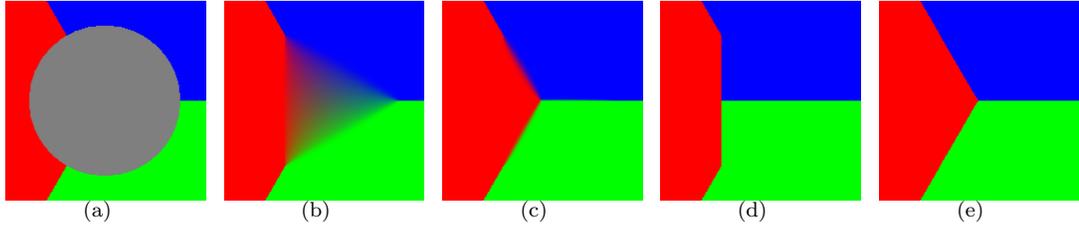


Figure 8: (a) Input, (b) Lellmann et. al., (c) Pock et. al., (d) Alpha expansion (e) dual model

Figure 9 (b) shows the result of 4 class segmentation of a brain MRI image. One would like to classify the input image in Figure 9 (a) into the classes: background, cerebrospinal fluid, gray matter and white matter. For this example we have used the Mumford-Shah model with  $L^2$  data term

$$f_i = |I - c_i|^2, \quad i = 1, \dots, 4.$$

In order to estimate the optimal constant values  $\{c_i\}_{i=1}^n$ , we alternate optimization with respect to  $\{c_i\}_{i=1}^n$  and the labeling function as described in more details in [1]. This algorithm finds a local minimum with respect the constant values. For energy plots, see Figure 10.

The positive parameter  $s$  controls how well the dual model is approximated. The lower  $s$  is the better the dual model is approximated. We found that setting  $s = 0.01$  or  $s = 0.005$  is sufficient and often optimal: setting  $s$  lower does not seem to lower the energy of the binary result. This indicates there is a certain benefit of the smoothing in connection with the thresholding scheme. This benefit can also be observed in the energy plots of Figure 10: we can obtain binary solutions of lower energy than the approaches of [45, 25].

## 5.2 Evaluation of efficiency and convergence

We now wish to compare the cpu time and convergence with the approaches [45, 25]. In order to deal with the simplex constraint in the primal optimization problem (5), an alternating optimization approach was used in [45] where one solve for  $k = 1, \dots$

$$u^{k+1} = \arg \min_u E_1(u) = \sum_{i=1}^n \int_{\Omega} \frac{1}{2\theta} (u_i - v_i^k)^2 + |\nabla u_i| dx \quad (32)$$

$$v^{k+1} = \arg \min_{v \in \Delta_+} E_2(v) = \sum_{i=1}^n \int_{\Omega} \frac{1}{2\theta} (u_i^{k+1} - v_i)^2 + \frac{1}{\lambda} f_i v_i dx, \quad (33)$$

Here  $\theta$  is some large parameter. The second problem can be optimized pointwise and has a closed form solution. However, the first subproblem is a TV optimiza-

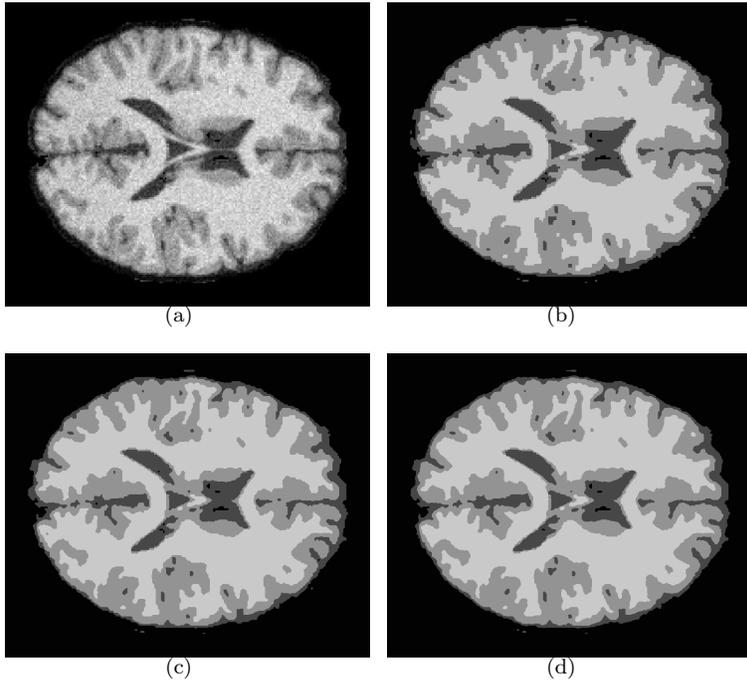


Figure 9: (a) Input, (b) alpha expansion 4 neighbors, (c) alpha expansion 8 neighbors, (d) dual model. Size:  $709 \times 591$ .

Table 1: Percentage of misclassified pixels for experiment 1-4

	$\alpha$ -expansion	$\alpha - \beta$ -swap	Lellmann et. al.	dual
Experiment1	8.89	6.12	-	5.51
Experiment2	1.17	1.17	-	1.06
Experiment3	7.42	15.72	12.30	11.72
Experiment4	6.64	7.23	6.25	5.86

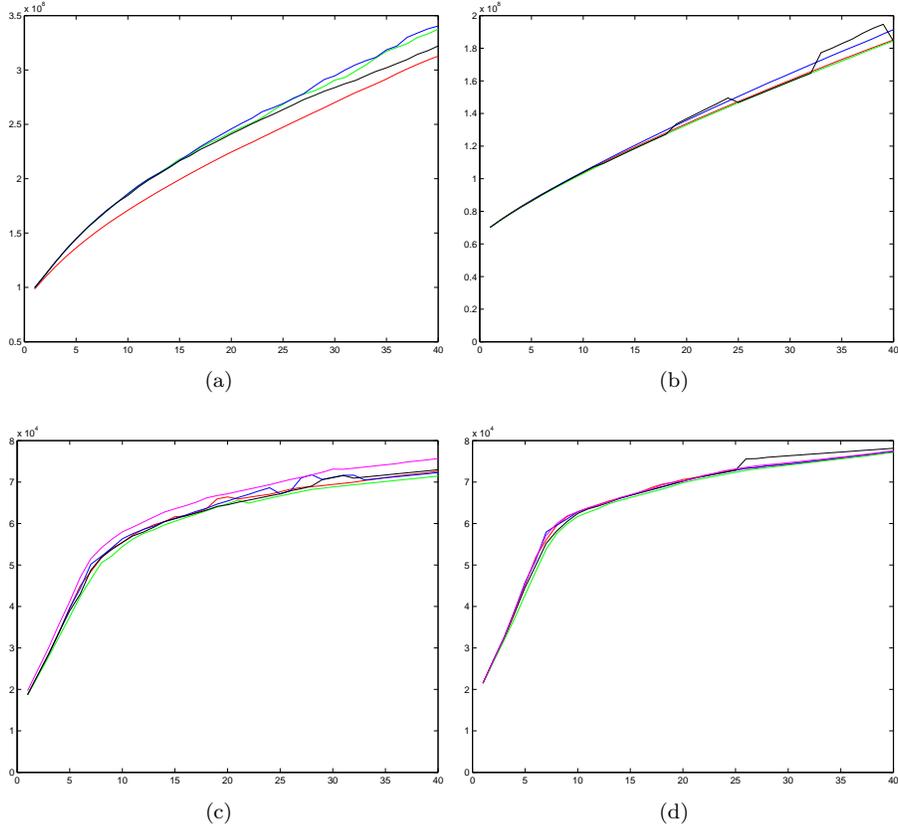


Figure 10: Energy plot as a function of regularization parameter  $\lambda$ . Red: smoothed dual model, green: alpha expansion 8 neighbors, blue: alpha expansion 4 neighbors, black: alpha-beta swap 8 neighbors, magenta: Zach et. al. [45]. (a) Flower, (b) brain (c)-(d) Experiment 3 and 4. In all experiments the smoothed dual model (red) performs better than or as good as competitive approaches. Fig (a) is a typically difficult example with a large number of labels, where the smoothed dual model clearly performs best. There are some bias in favor of the discrete models, since the final energy is eventually evaluated on a discrete grid.

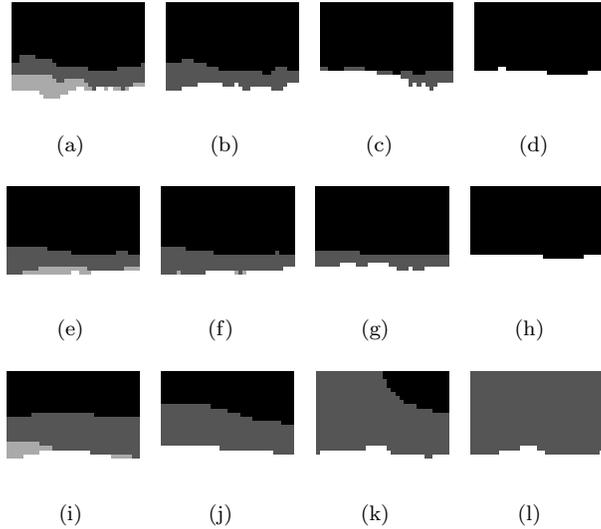


Figure 11: Convergence comparison on input image from Figure 6 (a) with highest regularization coefficient  $\lambda$  from Figure 10(c). Top row: Zach et. al. [45] with  $\theta = 0.01$ . (a) 500 outer iterations, (b) 1500 outer iterations (c) 3000 outer iterations, (d) 7000 outer iterations. Middle row: Douglas Rachford splitting from Lellmann et. al. [25] with outer time step size of 0.02. (e) 500 outer iterations, (f) 1000 outer iterations, (g) 2000 outer iterations, (h) 3500 outer iterations. Bottom row: smoothed dual model: (i) 500 iterations, (j) 1000 iterations, (k) 1500 iterations, (l) 2200 iterations. Within each outer iteration of [45, 25] a TV minimization problem must be solved iteratively.

tion problem, and must be solved by some iterative technique such as Chambolle’s algorithm [9]. In [25] a very similar scheme based on Douglas-Rachford splitting was used, which has proven convergence provided the subproblems are solved exactly. This scheme also involves a substep where a TV minimization problem needs to be solved iteratively. Therefore, both these schemes require one outer loop and one inner loop. In contrast, the simplex constraint is inherent in our dual formulation, therefore only one loop is enough. Furthermore, each iteration of this loop has a computational cost approximately equal to one inner loop iteration of [9, 25]. In order to make a reasonable comparison of efficiency, the subproblems in [45, 25] are assumed to be solved with limited accuracy. Even in those cases, we found our approach to be significantly more efficient. When  $\theta$  is low, the problem is solved with high accuracy, but more iterations are required. Therefore, one could say  $\theta$  plays the same role as the smoothing parameter  $s$  in our approach. When  $s$  is low, the relaxed problem is solved with higher accuracy, but more iterations are required as the time step size  $\delta$  depends on  $s$  to have stability. Trial and error indicate that this dependency is given by  $\delta \leq \frac{1}{2}s$  when images are scaled between 0 and 1. In practice we experienced that the number of outer iterations in [45, 25] exceeds the total number of iterations in our approach, especially when the regularization parameter  $\lambda$  is high. The Douglas-Rachford splitting approach of [45] tends to require an amount of outer iterations approximately equal to the total number of iterations in our approach when the subproblems are solved exactly. An illustrative example is given in Figure 11. Here we have tried to optimize all parameters, to minimize the number of outer iterations. Furthermore, each subproblem in [45, 25] is solved with high accuracy. Zach et. al. [45] needs 7000 outer iterations for this difficult example. Lellmann et. al. needs 3500 outer iterations, while our approach converges in a total of 2200 iterations. The final results are a little different, indicating that there is a difference between the thresholding schemes. As seen in the energy plot of Figure 10(c), the solution produced by our method has a lower energy.

Comparing the highly optimized C++ implementation [6] of alpha expansion and our matlab implementation of the dual model, alpha expansion was faster for all experiments as expected. However, there has recently been much effort on comparing continuous and discrete (graph cut) techniques for two phase partitioning problems [19]. Continuous techniques are getting close because of potential of parallel and GPU implementation. Unfortunately, we don’t have the resources to implement our method in such a highly optimized manner. We can mention that Zach et. al. implemented their algorithm on gpu and thereby claimed to beat the graph cut based approaches in terms of efficiency by a factor of 30. The cpu times for our matlab implementation are as follows: For the  $709 \times 591$  brain image convergence was reached in 1 minutes and 32 seconds for our implementation. For the  $32 \times 32$  images in Figure 6 and 7 convergence averaged around 2.5 seconds. For the  $100 \times 100$  images in Figure 4 and 5 convergence took 10.21 and 4.68 seconds respectively. Due to the extreme amount of noise on these small images the regularization parameter must be set very high, which increases cpu time compared to images of the same size with

lower noise level. In [25, 45] there are much more parameters to tune, like  $\theta$ , step sizes in algorithm for subproblem, accuracy of solving subproblems etc. We have done our best to optimize these parameters to make convergence as fast as possible. Even then we observed that our matlab implementations of these approaches generally require as least 20-30 times as much time to reach acceptable solutions near convergence.

## 6 Conclusions and Future Topics

This paper proposed a novel duality-based approach for the continuous multi-labeling problem, i.e. the convex relaxed Potts model. Through the dual formulation, we could explain why optimal solutions are often expected to be binary. Moreover, we suggested a smoothing method based on the log-sum exponential function, so as to deal with the nonsmooth dual problem, and also indicated that such smoothing approach formally gives rise to the novel smoothed primal-dual model and suggested labelings with maximum entropy. By this, the close connections between optimal labelings and geometrical clustering of spatial points were revealed. The numerical experiments showed that such smoothed method for the dual model produces an expectation maximization like algorithm for the multi-labeling problem and yields better numerical results. We will apply our current work for more imaging and vision tasks in future.

## A Proof of Prop. 4

Before we prove Prop. 4, we first give the following result

**Proposition 10.** Given a bounded scalar field  $u$  defined on  $\Omega$ , we assume, without loss of generality,  $0 \leq u(x) \leq 1$  for all  $x \in \Omega$ . If a vector field  $p^*$  maximizes the integral  $\int_{\Omega} u \operatorname{div} p \, dx$  over the convex set

$$C_{\lambda} := \{p \mid |p(x)| \leq \lambda, p_n = 0\}, \quad (34)$$

then for almost every level set  $u^{\gamma}$  of  $u$

$$u^{\gamma} = \begin{cases} 1, & \text{when } u(x) \geq \gamma \\ 0, & \text{when } u(x) < \gamma \end{cases}$$

with  $\gamma \in [0, 1]$ ,  $p^*$  also maximizes the same integral  $\int_{\Omega} u^{\gamma} \operatorname{div} p \, dx$  over the convex set  $C^{\lambda}$  and equals to  $\lambda L^{\gamma}$  where  $L^{\gamma}$  is the perimeter of the level set  $u^{\gamma}$ .

*Proof.* Denote the interval  $\Gamma = [0, 1]$ . The coarea formula is a powerful tool which says that

$$\int_{\Omega} |\nabla u| \, dx = \int_{\Gamma} \int_{\Omega} |\nabla u^{\gamma}| \, dx \, d\gamma. \quad (35)$$

By applying this formula we can deduce

$$\int_{\Omega} u \operatorname{div} p^* dx = \int_{\Omega} |\nabla u| dx = \int_{\Gamma} \int_{\Omega} |\nabla u^{\gamma}| dx d\gamma = \int_{\Gamma} \left( \max_{p \in C_{\lambda}} \int_{\Omega} u^{\gamma} \operatorname{div} p dx \right) d\gamma. \quad (36)$$

By the fact that  $u(x) = \int_0^{u(x)} d\gamma = \int_{\Gamma} u^{\gamma}(x) d\gamma$  for any  $x \in \Omega$ , we have

$$\int_{\Omega} u \operatorname{div} p^* dx = \int_{\Omega} \left( \int_{\Gamma} u^{\gamma}(x) d\gamma \right) \operatorname{div} p^*(x) dx = \int_{\Gamma} \int_{\Omega} u^{\gamma} \operatorname{div} p^* dx d\gamma. \quad (37)$$

Therefore, combining (36) and (37)

$$\int_{\Gamma} \int_{\Omega} u^{\gamma} \operatorname{div} p^* dx d\gamma = \int_{\Gamma} \left( \max_{p \in C_{\lambda}} \int_{\Omega} u^{\gamma} \operatorname{div} p dx \right) d\gamma. \quad (38)$$

This equality (38) together with the fact that for any  $\gamma \in [0, 1]$

$$\int_{\Omega} u^{\gamma} \operatorname{div} p^* dx \leq \max_{p \in C_{\lambda}} \int_{\Omega} u^{\gamma} \operatorname{div} p dx \quad (39)$$

implies that

$$\int_{\Omega} u^{\gamma} \operatorname{div} p^* dx = \max_{p \in C_{\lambda}} \int_{\Omega} u^{\gamma} \operatorname{div} p dx$$

for almost every  $\gamma \in [0, 1]$ . Clearly, the perimeter of the level set  $u^{\gamma}$  is given by

$$L^{\gamma} = \int_{\Omega} |\nabla u^{\gamma}| dx = \max_{p \in C_{\lambda}} \int_{\Omega} u^{\gamma} \operatorname{div} p dx.$$

□

Now we give the proof of Prop. 4.

*Proof.* Let  $u^d(x) = (u_1^d(x), \dots, u_n^d(x))$  be an optimum of the primal problem consistent with  $f_i + \operatorname{div} p_i^*$   $i = 1, \dots, n$ , as suggested in Theorem 1, i.e.

$$\forall x \in \tilde{\Omega}, \quad u_k^d(x) + u_j^d(x) = 1, \quad u_i^d(x) = 0, \quad i \neq k, j$$

and

$$E^D(p^*) = E(u^d, p^*) = E^P(u^d). \quad (40)$$

Obviously,  $p_i^*$  maximizes the integral

$$\int_{\Omega} u_i^d(x) \operatorname{div} p_i^*(x) dx, \quad i = 1, \dots, n$$

over the convex set  $C_{\lambda}$  (34).

Now we define  $\tilde{u}(x) = (\tilde{u}_1(x), \dots, \tilde{u}_n(x))$  as follows: let

$$\tilde{u}_i(x) = u_i^d(x), \quad x \in \Omega, \quad i \neq j, k$$

which are obviously binary as  $u_i^d(x) \in \{0, 1\}$ ,  $i \neq j, k$ , by virtue of Theorem 1.

For  $\tilde{u}_j(x)$  and  $\tilde{u}_k(x)$ , we choose any value  $\gamma \in (0, 1)$  and let

$$\tilde{u}_k(x) = \begin{cases} 1, & \text{when } u_k^d(x) \geq \gamma \\ 0, & \text{when } u_k^d(x) < \gamma \end{cases}, \quad x \in \tilde{\Omega}$$

and

$$\tilde{u}_k(x) = u_k^d(x), \quad x \in \Omega \setminus \tilde{\Omega}.$$

It is easy to see that  $\tilde{u}_k(x)$  is binary in this setting. By such a configuration of  $\tilde{u}_k$ , let

$$\tilde{u}_j(x) = 1 - \tilde{u}_k(x), \quad x \in \tilde{\Omega}$$

and

$$\tilde{u}_j(x) = u_j^d(x), \quad x \in \Omega \setminus \tilde{\Omega},$$

which is also binary in  $\Omega$ .

It is easy to see that the above binary setting of  $\tilde{u}_i(x)$ ,  $i = 1, \dots, n$  doesn't change the total energy of (14), i.e.

$$E(\tilde{u}, p^*) = \sum_{i=1}^n \int_{\Omega} \tilde{u}_i(x) (f_i(x) + \operatorname{div} p_i^*(x)) dx = E^D(p^*). \quad (41)$$

Moreover, by Prop. 10, we can easily verify that  $p_k^*$  maximizes the integral  $\int_{\Omega} \tilde{u}_k \operatorname{div} p_k dx$  over the convex set  $C_{\lambda}$ . By the fact

$$u_j^d(x) = 1 - u_k^d(x), \quad \tilde{u}_j(x) = 1 - \tilde{u}_k(x), \quad x \in \tilde{\Omega}$$

and  $u_j^d(x), \tilde{u}_j(x) \in \{0, 1\}$  when  $x \in \Omega \setminus \tilde{\Omega}$ , then  $\tilde{u}_j(x)$  is the  $1 - \gamma$  level set of  $u_j^d(x)$ . Therefore  $p_j^*$  also maximizes the integral  $\int_{\Omega} \tilde{u}_j \operatorname{div} p_k dx$  over the convex set  $C_{\lambda}$ .

By the above facts, the total energy (41) related to  $\tilde{u}_i$  is actually  $E^P(\tilde{u})$ , hence we have

$$E^P(\tilde{u}) = E(\tilde{u}, p^*) = E^D(p^*).$$

It follows that  $\tilde{u}$  is a minimum of the primal problem (5) which is also binary.  $\square$

## B Proof of Prop. 9

Proof of Prop. 9 is shown as follows

*Proof.* Let  $u_i^s$ ,  $i = 1, \dots, n$ , be given by (25), i.e.  $u^s$  solves the smoothed primal model (23), see Prop. 8.

At the position  $x \in \Omega$ , let  $f_k(x) + \operatorname{div} p_k^s(x)$  and  $f_l(x) + \operatorname{div} p_l^s(x)$  be the first and second minimizer of  $f_i(x) + \operatorname{div} p_i^s(x)$ ,  $i = 1, \dots, n$ , i.e. we have

$$f_k(x) + \operatorname{div} p_k^s(x) < f_l(x) + \operatorname{div} p_l^s(x) \leq f_i(x) + \operatorname{div} p_i^s(x), \quad i \neq k, l,$$

as

$$d^s(x) = (f_l(x) + \operatorname{div} p_l^s(x)) - (f_k(x) + \operatorname{div} p_k^s(x)) > 0$$

in view of  $D^s = \min_{x \in \Omega} d^s(x) > 0$ .

Given

$$s \leq \frac{d^s(x)}{\log [\gamma(n-1)/(1-\gamma)]}, \quad (42)$$

where  $\gamma \in [0.5, 1)$ , then

$$\exp\left(\frac{d^s(x)}{s}\right) \geq \frac{\gamma(n-1)}{1-\gamma}.$$

It follows that

$$\frac{1}{1 + (n-1) \exp\left(-\frac{d^s(x)}{s}\right)} \geq \gamma. \quad (43)$$

By (25),

$$u_k^s(x) \geq \frac{\exp\left(\frac{-f_k(x) - \operatorname{div} p_k^s(x)}{s}\right)}{\exp\left(\frac{-f_k(x) - \operatorname{div} p_k^s(x)}{s}\right) + (n-1) \exp\left(\frac{-f_l(x) - \operatorname{div} p_l^s(x)}{s}\right)}.$$

Therefore, by (43) we have

$$u_k^s(x) \geq \gamma,$$

which means that for properly chosen  $s$  by (42)  $u_k^s(x) \geq \gamma$  at each  $x \in \Omega$ .

Therefore, when

$$s \leq \frac{D^s}{\log [\gamma(n-1)/(1-\gamma)]},$$

for each position  $x \in \Omega$

$$u_k^s(x) \geq \gamma,$$

where  $k$  indicates the component such that  $f_k(x) + \operatorname{div} p_k^s(x)$  is the minimizer of  $f_i(x) + \operatorname{div} p_i^s(x)$   $i = 1, \dots, n$ .

Moreover, by (25), we have

$$\sum_{i=1}^n u_i^s(x) = 1, \quad \text{and} \quad u_i^s(x) \geq 0, \quad i = 1, \dots, n$$

for each  $x \in \Omega$ . Therefore, for  $u_k^s(x) \geq \gamma$  and  $\gamma \in [0.5, 1)$

$$\sum_{i \neq k} u_i^s(x) \leq 1 - \gamma \leq 0.5.$$

Let  $(u_i^s)^\gamma$  be the  $\gamma$  upper level-set function of  $u_i^s$  such that

$$(u_i^s)^\gamma(x) := \begin{cases} 1 & \text{if } u_i^s(x) \geq \gamma \\ 0 & \text{otherwise} \end{cases}, \quad x \in \Omega.$$

Hence, at each  $x \in \Omega$  we have

$$(u_k^s)^\gamma = 1, \quad (u_i^s)^\gamma(x) = 0, \quad i \neq k,$$

Obviously the thresholded result  $\tilde{u}^s$  by (27) can also be achieved by the  $\gamma$  upper level set function  $(u_i^s)^\gamma$ , i.e.

$$\tilde{u}_i^s = (u_i^s)^\gamma, \quad i = 1, \dots, n. \quad (44)$$

In view of the smoothed primal-dual model (22) and the smoothed primal model (23), we see that the dual functions  $p_i^s$ ,  $i = 1, \dots, n$ , just maximize the integrals

$$\int_{\Omega} u_i^s(x) \operatorname{div} p_i^s(x) dx, \quad i = 1, \dots, n$$

over the convex set  $C_\lambda$ . Then by Prop. 10,  $p_i^s$ ,  $i = 1, \dots, n$ , maximize

$$\int_{\Omega} (u_i^s)^\gamma(x) \operatorname{div} p_i^s(x) dx, \quad i = 1, \dots, n,$$

over  $C_\lambda$ . It follows, by (44),  $p_i^s$ ,  $i = 1, \dots, n$ , also maximize

$$\int_{\Omega} \tilde{u}_i^s(x) \operatorname{div} p_i^s(x) dx, \quad i = 1, \dots, n$$

over  $C_\lambda$ .

In this regard,

$$E^P(\tilde{u}^s) = E(\tilde{u}^s, p^s) = E^D(p^s).$$

It directly indicates that the thresholded result  $\tilde{u}^s$  is a global solution to the nonsmooth primal model (5). As  $\tilde{u}_i^s$ ,  $i = 1, \dots, n$ , are also binary,  $\tilde{u}^s$  is also a solution to the nonconvex Potts model (4).  $\square$

## References

- [1] E. Bae and X.C. Tai. Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. In *Scale Space and Variational Methods in Computer Vision*, pages 1–13. Springer, 2009.
- [2] Egil Bae and Xue-Cheng Tai. Efficient global minimization for the multiphase Chan-Vese model of image segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 28–41, 2009.
- [3] Arindam Banerjee, Srujana Merugu, Inderjit Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. In *Journal of Machine Learning Research*, pages 234–245, 2004.

- [4] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*, pages 26–33, Washington, DC, USA, 2003. IEEE Computer Society.
- [5] Yuri Boykov, Vladimir Kolmogorov, Daniel Cremers, and Andrew Delong. An integral solution to surface evolution pdes via geo-cuts. In *In ECCV*, pages 409–422, 2006.
- [6] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1222 – 1239, 2001.
- [7] X. Bresson, S. Esedoglu, P. Vandergheynst, J.P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.
- [8] Ethan S. Brown, Tony F. Chan, and Xavier Bresson. Convex formulation and exact global solutions for multi-phase piecewise constant Mumford-Shah image segmentation. *UCLA, Applied Mathematics, CAM-report-09-66*, July, 2009.
- [9] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, January 2004.
- [10] T. Chan and L.A. Vese. Active contours without edges. 10:266–277, 2001.
- [11] Ivar Ekeland and Roger Téman. *Convex analysis and variational problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.
- [12] Ky Fan. Minimax theorems. *Proc. Nat. Acad. Sci. U. S. A.*, 39:42–47, 1953.
- [13] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, pages 271–279, 1989.
- [14] J. M. Hyman and M. J. Shashkov. Adjoint operators for the natural discretizations of the divergence, gradient and curl on logically rectangular grids. *Appl. Numer. Math.*, 25(4):413–442, 1997.
- [15] J. M. Hyman and M. J. Shashkov. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Comput. Math. Appl.*, 33(4):81–104, 1997.
- [16] Hiroshi Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:1333–1336, 2003.

- [17] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [18] Krzysztof C. Kiwił. Proximal minimization methods with generalized bregman functions. *SIAM Journal on Control and Optimization*, 35:1142–1168, 1995.
- [19] Maria Klodt, Thomas Schoenemann, Kalin Kolev, Marek Schikora, and Daniel Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 332–345, Berlin, Heidelberg, 2008. Springer-Verlag.
- [20] Pushmeet Kohli, M. Pawan Kumar, and Philip H.S. Torr.  $p^3$  and beyond: Move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1645–1656, 2009.
- [21] Vladimir Kolmogorov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *ICCV*, pages 564–571, 2005.
- [22] Vladimir Kolmogorov and Ramin Zabih. Multi-camera scene reconstruction via graph cuts. In *European Conference on Computer Vision*, pages 82–96, 2002.
- [23] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:65–81, 2004.
- [24] Nikos Komodakis and Georgios Tziritas. Approximate labeling via graph-cuts based on linear programming. In *Pattern Analysis and Machine Intelligence*, pages 1436–1453, 2007.
- [25] Jan Lellmann, Jörg Kappes, Jing Yuan, Florian Becker, and Christoph Schnörr. Convex multi-class image labeling by simplex-constrained total variation. In *SSVM '09: Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision*, pages 150–162, Berlin, Heidelberg, 2009. Springer-Verlag.
- [26] H. Li and X.C. Tai. Piecewise constant level set methods for multiphase motion. *International Journal of Numerical Analysis and Modeling*, 4:274–293, 2007.
- [27] J. Lie, M. Lysaker, and X.-C. Tai. A variant of the level set method and applications to image segmentation. *Math. Comp.*, 75(255):1155–1174, 2006.

- [28] J. Lie, M. Lysaker, and X.C. Tai. A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE Transactions on Image Processing*, 15(5):1171–1181, 2006.
- [29] Yves Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.
- [30] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42, 42:577–685, 1989.
- [31] Mila Nikolova, Selim Esedoglu, and Tony F. Chan. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.
- [32] S. Osher and J.A. Sethian. Fronts propagating with curvature dependent speed: algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.
- [33] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *European Conference on Computer Vision (ECCV)*, Marseille, France, October 2008.
- [34] Thomas Pock, Thomas Schoenemann, Gottfried Graber, Horst Bischof, and Daniel Cremers. A convex formulation of continuous multi-label problems. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 792–805, Berlin, Heidelberg, 2008. Springer-Verlag.
- [35] Renfrey B. Potts. Some generalized order-disorder transformations. In *Proceedings of the Cambridge Philosophical Society, Vol. 48*, pages 106–109, 1952.
- [36] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
- [37] Kenneth Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. In *Proceedings of the IEEE*, pages 2210–2239, 1998.
- [38] Gilbert Strang. Maximal flow through a domain. *Mathematical Programming*, 26:123–143, 1983.
- [39] W. Tao and X. Tai. Multiple Piecewise Constant Active Contours for Image Segmentation using Graph Cuts Optimization. *UCLA cam-report 09-13*, 2009.
- [40] Marc Teboulle. A unified continuous optimization framework for center-based clustering methods. *J. Mach. Learn. Res.*, 8:65–102, 2007.

- [41] Olga Veksler. Image segmentation by nested cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 339–344, 2000.
- [42] L. A. Vese and T. F. Chan. A new multiphase level set framework for image segmentation via the mumford and shah model. *International Journal of Computer Vision*, 50:271–293, 2002.
- [43] Martin Wainwright, Tommi Jaakkola, and Alan Willsky. Map estimation via agreement on (hyper)trees: Message-passing and linear programming approaches. *IEEE Transactions on Information Theory*, 51:3697–3717, 2002.
- [44] J. Yuan, C. Schnörr, and G. Steidl. Simultaneous optical flow estimation and decomposition. *SIAM J. Scientific Computing*, 29(6):2283–2304, 2007.
- [45] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization Workshop (VMV)*, 2008.