Global Minimization for Continuous Multiphase Partitioning Problems Using a Dual Approach

Egil Bae · Jing Yuan · Xue-Cheng Tai

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 tackling 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 show 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 develop a smoothing method based on the log-sum exponential function and indicate that such a smoothing approach leads to a novel smoothed primal-dual model and suggests labelings with maximum entropy. Such a smoothing method for the dual model also yields a new thresholding scheme to obtain approximate solutions. An expectation maximization like algorithm is proposed based on the smoothed formulation which is shown to be superior in efficiency compared to earlier approaches from continuous optimization. Numerical experiments also show that our method outperforms several competitive approaches in various aspects, such as lower energies and better visual quality.

Keywords Convex relaxation · Image segmentation · Primal-dual methods · Total variation

E. Bae (✉) · X.-C. Tai
Department of Mathematics, University of Bergen, Bergen, Norway
e-mail: Egil.Bae@math.uib.no

X.-C. Tai
e-mail: tai@math.uib.no

J. Yuan
Department of Computer Science, University of Western Ontario, London, Canada
e-mail: cn.yuanjing@googlemail.com

X.-C. Tai
Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore

1 Introduction

The multiphase partitioning problem, or multi-labeling problem, is the problem of partitioning the image domain into several regions according to some optimization criterion. The term "multiphase problem" is often used in the variational and PDE community. In the discrete optimization community, such problems are often called "multi-labeling problems" and are expressed as the assignment of a label l at each point of the image domain chosen from a finite set of real numbers $\{l_1, \ldots, l_n\}$. Such problems appear extensively in the areas of image processing and computer vision. They are typically formulated as the minimization of an energy function which mathematically encodes all the information needed for the imaging and vision task. Lower energy solutions are regarded as better posteriori estimates. 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. The Potts model can, for instance, model image segmentation problems where one wants to assign each pixel to a region by minimizing a weighted sum of data fidelity and the lengths of the boundaries of the regions. The piecewise constant

Mumford-Shah model (Mumford and Shah 1989) is a special case of the Potts model, if one assumes the "constants" are fixed.

One can solve such labeling problems in the manner of "discrete" or "continuous", depending on the spatial definition of l. When the labeling is defined on a discrete grid, it reduces to a 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 (see Li 2001; Paragios et al. 2005 for a good reference). The node cost often evaluates the fidelity of the given data. The edge cost, often pairwise (Boykov et al. 2001; Kolmogorov and Zabih 2004) or with high-order cliques (Kohli et al. 2009), measures the regularities of the label assignments. In case of two labels such energy functions can be efficiently and globally minimized by graph cuts (Greig et al. 1989), provided they are submodular (Kolmogorov and Zabih 2004). Exact global minimization is also possible in certain cases with multiple labels (Ishikawa 2003; Bae and Tai 2009b). However, most problems with more than two labels are NP-hard in the discrete context. Therefore they can probably not be solved globally in polynomial time. The Potts model is an important example of such a NP-hard problem. Several approximation algorithms for the Potts model exist, e.g. message passing (Wainwright et al. 2002; Kolmogorov 2006), linear programming (Komodakis and Tziritas 2007) etc. Most notable is the graph cuts based alpha expansion and alpha-beta swap (Boykov et al. 2001). Despite the efficiencies of these discrete approaches, their computation results are biased by the discrete grid, therefore metrication errors are introduced. By considering more neighboring nodes, such visual artifacts can be reduced (Boykov et al. 2006; Kolmogorov 2005). However, this results in both a heavier memory load and a larger computational cost. Figure 1 shows an example where the image in (a) has been segmented into 10 regions by minimizing the Potts energy. An approximate solution which is generated by graph cut based alpha expansion implemented with an "8 neighborhood system" is shown in Fig. 1(b). In this example, each color represents a distinct region. The Potts model seeks to minimize the length the region boundaries.
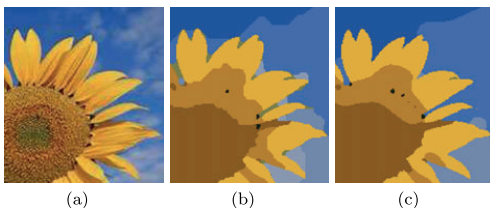

Fig. 1 Segmentation into 10 regions by using Potts model. (**a**) Input, (**b**) result graph cut-based alpha-expansion (Boykov et al. 2001) with "8-neighborhood system", (**c**) result our approach. The full image and experiment is presented in Sect. 5

It is seen that horizontal, vertical and diagonal orientations of the boundaries are favored. In addition, some artifacts arise, because the energy is not minimized exactly: observe the green patches along the intersection between the yellow flower and blue sky.

On the other hand, the labeling problem can also be formulated in the continuous setting, where the continuous domain Ω is divided into two or more subregions $\{\Omega_i\}_{i=1}^n$ by minimizing certain energy functionals. To find a solution numerically, the descent or ascent equations are discretized, instead of solving a discrete problem directly. Such continuous formulations are not biased by the grid, and can therefore produce results that are rotationally invariant. For instance, the Euclidean length of the boundaries of Ω_i can be represented without considering infinitely large neighborhood systems. The level set method (Osher and Sethian 1988) is an elegant tool for simplifying these energy functionals, such that numerical calculation can be easily implemented (Chan and Vese 2001; Vese and Chan 2002). Fairly general energy functionals, such as the continuous variant of the Potts model, can be rewritten and minimized numerically by level sets. A variant of the level set method, the so-called piecewise constant level set method (PCLSM) (Lie et al. 2006a, 2006b), expresses the energy in terms of a labeling function, or piecewise constant function as the authors called it. Unfortunately, the main disadvantage of both level set methods is their potential of getting stuck in a possibly inferior local minima because of their nonconvex formulations.

Recently, it was realized that certain binary labeling problems in the spatially continuous setting can be solved in a global fashion (Nikolova et al. 2006; Chan and Esedoglu 2005). The non-convex binary constraint of the labeling function can be relaxed such that the minimization problem becomes convex. It was proved that thresholding the solution of the relaxed problem at any value in [0, 1] yields a globally optimal solution to the original binary labeling problem. Therefore fast and reliable algorithms based on convex optimization theories can be applied, which have been shown to be much faster than the level set method, see e.g. (Bresson et al. 2007). Specialized implementations can even be more efficient than graph cuts (Goldstein et al. 2009; Klodt et al. 2008). Moreover, such convex optimization algorithms consists mainly of floating point matrix and vector arithmetic and is therefore highly suited for massive parallel implementation on cpu. Combinatorial algorithms for graph cuts have a much more serial nature which makes them difficult to implement on GPU.

Recently, similar convex formulations have also been proposed for the continuous Potts model with more than two labels (Zach et al. 2008; Lellmann et al. 2009; Pock et al. 2009), by relaxing the integrability constraint of the labeling function. However, no proof exist that integer valued

solutions exist to these relaxations. One can obtain approximate integer-valued solutions by simple thresholdings. Experiments show such approximations are much closer to the exact global minimum than the local minima computed by level sets. Some other relaxations closely related to (Zach et al. 2008; Lellmann et al. 2009) have later been proposed in (Brown et al. 2009, 2010).

1.1 Contributions and Organization

This paper builds on the work of Zach et al. (2008) and Lellmann et al. (2009). We discuss the convex relaxed Potts model under a primal-dual perspective. A novel dual model is proposed, which formulates the convex relaxed problem in terms of the dual variables only and thereby avoids the complex side constraint of the primal variable. This contrasts with previous approaches which tackle the constrained optimization problem over the labeling function directly. The dual formulation also provides a new thresholding scheme for obtaining binary primal solutions of the relaxed problem which are also globally optimal to Potts model, and an analytical tool to derive sufficient conditions about when this is possible, i.e. for when the relaxation is exact. In order to deal with non-smoothness of the resulting dual energy functional, a smoothed version of the convex dual model is proposed. Such a smoothing scheme for the dual model has some interesting connections to the formulations of maximum entropy clustering algorithms (Rose 1998). A new algorithm derived from the smoothed dual formulation is shown to be significantly more efficient than the state of art works (Zach et al. 2008; Lellmann et al. 2009). It is very simple and can be easily implemented. Experiments demonstrate the algorithm for the smoothed model may yield better binary approximations to the original non-convex problem with lower energy in connection with the new thresholding scheme. Numerical results are also compared extensively with the well-known algorithms alpha expansion and alpha-beta swap from discrete optimization, which show that our algorithm can produce results of equal or lower energy than these approaches. In addition our algorithm avoids the grid bias and yields results that are rotationally invariant. For an example, see our result on the flower image in Fig. 1(c). In contrast to graph cut based alpha expansion, the orientation of the boundaries are not biased to certain direction and have a much smoother and more natural appearance. In addition, artifacts, like the green misclassifications along the transition between flower and sky, are avoided. Further experiments show that the result produced by our approach has a lower energy for this example. In contrast to graph cuts, our algorithm is also very parallel friendly and can be easily implemented on GPU.

This paper is organized as follows: In Sect. 2 we introduce the continuous Potts model and its convex relaxation.

Then we propose the primal, primal-dual and dual formulations of the convex relaxed Potts model, which are equivalent to each other. Analysis of the dual formulation are given in Sect. 2.3 where sufficient conditions are derived for when binary primal solutions can be recovered from a dual solution. The smoothed models are presented in Sect. 3, Sect. 3.2 reveals a connection to maximum entropy labeling algorithms and error bounds of the smooth models are derived in Sect. 3.3. The new algorithm presented in Sect. 4, along with details on numerical implementation. Numerical experiments and comparisons are presented in Sect. 5.

The sections are arranged in increasingly advanced order. The most practically minded reader may find the first half of each section most interesting and can alternatively follow the order: Sects. 2 until (but excluding) 2.3, Sects. 3 until (but excluding) 3.2, Sect. 4 and Sect. 5.

2 Continuous Potts Model and a Convex Relaxation Approach

This work builds on the convex relaxation method of the nonconvex Potts model (Zach et al. 2008; Lellmann et al. 2009). We develop a novel dual formulation of the relaxation which simplifies the problem and leads to simpler computation. The dual formulation also allows to give a deeper analysis of the relaxation. It is shown there exists optimal solutions that are closely related to global optimums of the original nonconvex Potts model, and which can be obtained in a new thresholding framework.

2.1 Continuous Potts Model

The continuous variant of the Potts model (Potts 1952) partitions the continuous domain Ω into n disjoint subdomains $\{\Omega_i\}_{i=1}^n$ by minimizing

$$\min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} f_i(x)dx + \alpha \sum_{i=1}^n |\partial\Omega_i|$$

$$\text{s.t.} \quad \bigcup_{i=1}^n \Omega_i = \Omega, \quad \Omega_k \cap \Omega_l = \emptyset, \forall k \neq l, \tag{1}$$

where $|\partial\Omega_i|$ measures the lengths of the boundaries of the disjoint subdomains Ω_i, $i = 1, \ldots, n$. The functions f_i, $i = 1, \ldots, n$, defined on Ω are given and evaluate the performance of label assignment at each partition Ω_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 labelings with "tight" or smooth boundaries. We will focus on image processing problems, in which Ω is simply the image domain in N-D (typically 2-D). In this case, the functions f_i, $i = 1, \ldots, n$, typically depend on the values of the input image I.

The Potts model has a close resemblance to the piecewise constant Mumford-Shah model (Mumford and Shah 1989)

$$\min_{\{\Omega_i\}_{i=1}^n, \mathbf{c}} \sum_{i=1}^n \int_{\Omega_i} |I - c_i|^\beta dx + \alpha \sum_{i=1}^n |\partial \Omega_i|$$

$$\text{s.t.} \quad \bigcup_{i=1}^n \Omega_i = \Omega, \quad \Omega_k \cap \Omega_l = \emptyset, \ \forall k \neq l, \tag{2}$$

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

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

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

The boundary lengths of the disjoint subdomains are given by

$$|\partial \Omega_i| = \int_\Omega |\nabla u_i| dx, \quad i = 1, \ldots, n. \tag{3}$$

The Potts model (1) can then be rewritten as

$$\min_{u_i(x) \in \{0,1\}} \sum_{i=1}^n \int_\Omega u_i(x) f_i(x) dx + \alpha \sum_{i=1}^n \int_\Omega |\nabla u_i| dx \tag{4}$$

subject to

$$\sum_{i=1}^n u_i(x) = 1 \quad \forall x \in \Omega.$$

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 formulations of the relaxed problem as a primal-dual model and a dual model. The dual formulation deals with the simplex constraint implicitly and provides a new framework for obtaining optimal binary solutions of the primal problem.

2.2.1 Primal Model

The binary constraints of (4) were relaxed in (Zach et al. 2008), by minimizing $u_i(x)$ over the interval [0, 1] instead. This gives rise to the convex relaxed Potts model defined as

$$\min_{u \in S} E^P(u) = \sum_{i=1}^n \int_\Omega u_i(x) f_i(x) dx + \alpha \sum_{i=1}^n \int_\Omega |\nabla u_i| dx \tag{5}$$

where the simplex constraint $\triangle_+$ is defined as:

$$\triangle_+ = \left\{ u \in \mathbb{R}^n \ \middle| \ \sum_{i=1}^n u_i = 1; \ u_i \geq 0, i = 1, \ldots, n \right\} \tag{6}$$

and the set S is

$$S = \{u : \Omega \mapsto \mathbb{R}^N | u(x) \in \triangle_+ \ \forall x \in \Omega\}.$$

If the minimizer of (5) happens to be binary everywhere, it is also a global minimizer of the original problem (4). However, unlike the two label problem, if the computed minimizer of (5) is not binary, there are no thresholding scheme which can convert to a binary global minimizer of (4). Even if such a binary minimizer exists, (5) may result in nonbinary solutions due to non-uniqueness. Zach et al. (2008) and Lellmann et al. (2009) proposed to use the indicator function of the largest component u_i as an approximate binary solution, i.e. the thresholded solution $\tilde{u}$ was selected as

$$\tilde{u}_k(x) = \begin{cases} 1 & \text{if } k = \arg \max_{i=1,\ldots,n} u_k^* , \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

If the above maximizer is not unique, the convention of using the maximizer with smallest index was suggested.

In this paper, we call the continuous optimization problem (5) *primal formulation* or *primal model* and u_i, $i = 1, \ldots, n$, *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 (Kass et al. 1988) are powerful for many image processing problems. It was shown in (Bresson et al. 2007; Tao and Tai 2009) 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

$$\min_{u_i(x) \in \{0,1\}} \sum_{i=1}^n \int_\Omega \{u_i(x) f_i(x) + \alpha \tilde{g}(x) |\nabla u_i|\} dx, \tag{8}$$

subject to $\sum_{i=1}^n u_i(x) = 1$. Likewise, its convex relaxed version is

$$\min_{u \in S} \sum_{i=1}^n \int_\Omega \{u_i(x) f_i(x) + \alpha \tilde{g}(x) |\nabla u_i|\} dx, \tag{9}$$

subject to (6). 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_σ 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, (8) reduces to the Potts model (4), and (9) corresponds to (5).

In this work, we discuss the convex relaxed Potts model (5) and its generalized model (9) 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

$$\alpha \int_\Omega |\nabla u| dx = \max_{p \in C_\alpha} - \int_\Omega \nabla u \cdot p \, dx = \max_{p \in C_\alpha} \int_\Omega u \operatorname{div} p \, dx \tag{10}$$

in terms of the dual variable p over the convex set C_α defined as

$$C_\alpha := \{p : \Omega \mapsto \mathbb{R}^N \mid |p(x)|_2 \le \alpha \,\, \forall x \in \Omega, \, p \cdot \mathbf{n}|_{\partial\Omega} = 0\}, \tag{11}$$

where $\mathbf{n}$ is the normal vector of $\partial\Omega$, see e.g. (Meyer 2001).

For the more general model (9), C_α is given by

$$C_\alpha := \{p : \Omega \mapsto \mathbb{R}^N \mid |p|_2 \le \tilde{g}, \, p \cdot \mathbf{n}|_{\partial\Omega} = 0\}. \tag{12}$$

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

$$\min_u \max_{p_i \in C_\alpha} E(u, p) = \sum_{i=1}^n \int_\Omega u_i(x) \big(f_i(x) + \operatorname{div} p_i(x)\big) dx. \tag{13}$$

The variables p_i $i = 1, \ldots, n$ are named as *dual variables* in this paper. Therefore, the min-max problem (13) 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 (13) can be interchanged

$$\min_{u \in S} \max_{p_i \in C_\alpha} E(u, p) = \max_{p_i \in C_\alpha} \min_{u \in S} E(u, p) \tag{14}$$

because the conditions of the minimax theorem (see e.g., Ekeland and Téman 1999, Chap. 6, Proposition 2.4, also Fan 1953) are all satisfied. That is, C_α and $\triangle_+$ are convex, and the energy function $E(u, p)$ is linear in 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 Ekeland and Téman (1999).

2.2.3 Dual Model

We will now derive another equivalent formulation of (5) by optimizing the primal-dual model (13) via the primal variable $(u_1(x), \ldots, u_n(x)) \in \triangle_+$ at each position $x \in \Omega$.

Observe that for any vector $q = (q_1, \ldots, q_n) \in \mathbb{R}^n$

$$\min_{(v_1, \ldots, v_n) \in \triangle_+} \sum_{i=1}^n v_i q_i = \min(q_1, \ldots, q_n), \tag{15}$$

Therefore, minimizing (13) over the primal variables $u_i(x)$ $i = 1, \ldots, n$, at each position $x \in \Omega$, leads to

$$\max_{p_i \in C_\alpha} E^D(p) := \int_\Omega \big\{ \min(f_1 + \operatorname{div} p_1, \ldots, f_n + \operatorname{div} p_n) \big\} dx. \tag{16}$$

We call (16) the *dual model* of (9). $E^D(p)$ is called the dual energy functional.

By regarding $d(l_i, x) = f_i(x) + \operatorname{div} p_i(x)$, $x \in \Omega$, as the proximity measure of labeling x as l_i, $i = 1, \ldots, n$, the minimal distance indicates which label should be assigned to x by the dual model (16). In this sense, the dual formulation (16) can be viewed as a generalized center-based clustering formulation (Teboulle 2007; Banerjee et al. 2004), 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 to a convex set. Moreover, updating $\operatorname{div} p_i(x)$ associated with (10) minimizes the perimeter of the spatial partitions Ω_i, $i = 1, \ldots, n$, implicitly! This gives a geometrical explanation of the dual model (16) in the sense of minimal length clusterings.

2.2.4 Discussions and Comments

Now we have two equivalent optimization models, the primal-dual model (13) and the dual model (16), 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 (13) over the dual variable p first, i.e.

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

Likewise, the energy functional $E^D(p)$ of the dual model (16) is resulted by minimizing $E(u, p)$ first over u, i.e.

$$E^D(p) := \min_{u \in S} E(u, p).$$

As a consequence, we always have

$$E^P(u) \ge E(u, p) \ge E^D(p). \tag{17}$$

Let (u^*, p^*) be optimal to the primal-dual model (13)), then by (14) we have

$$E^P(u^*) = E(u^*, p^*) = E^D(p^*). \qquad (18)$$

Pairs (u^*, p^*) that satisfies (18) are called optimal primal-dual pairs.

Moreover, the equivalences between these models implies that we can also solve the convex relaxed Potts problem (5) by optimizing its dual model (16). In fact, when the n values $(f_1(x) + \text{div } p_1^*(x), \ldots, f_n(x) + \text{div } p_n^*(x))$ 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 (15), by

$$u_k^*(x) = \begin{cases} 1 & \text{if } k = \arg\min_{i=1,\ldots,n}(f_i(x) + \text{div } p_i^*(x)), \\ 0 & \text{otherwise,} \end{cases} \qquad (19)$$

which is a binary indicator vector and exactly indicates the optimal label function u^* at position x. Such a 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, $i = 1, \ldots, n$, directly. Moreover, the dual model (16) 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 relationship between a maximum p^* of the dual model (16) and a minimum u^* of the primal model (5) as follows

Theorem 1 *Given any maximum p^* of the dual problem (16). We will characterize primal variables u^* for which (u^*, p^*) is an optimal primal-dual pair of (13). If the n values $(f_1(x) + \text{div } p_1^*(x), \ldots, f_n(x) + \text{div } p_n^*(x))$, at some $x \in \Omega$, have a unique minimum, e.g. $f_k(x) + \text{div } p_k^*(x)$, then $u^*(x)$ at x must be valued*

$$u_k^*(x) = 1 \quad \text{and} \quad u_i^*(x) = 0, i \neq k. \qquad (20)$$

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

$$\sum_{i=1}^{k} u_{t_i}^*(x) = 1 \quad \text{and} \quad u_j^*(x) = 0, \quad j \notin \{t_1, \ldots, t_k\}. \qquad (21)$$

Proof Assume there exists a primal variable u^* such that (u^*, p^*) is an optimal primal-dual pair and u^* does not satisfy (20) and (21) for all $x \in \Omega$. Let $x \in \Omega$ be a point where (20) or (21) are violated, then for some $\varepsilon \in (0, 1]$, $u^*(x)$ satisfies

$$\sum_{i \in \mathcal{T}} u_i^*(x) = 1 - \varepsilon, \qquad \sum_{i \in \{1,\ldots,n\} \setminus \mathcal{T}} u_i^*(x) = \varepsilon.$$

Let $\min_{i \in \{1,\ldots,n\}}^{2\text{nd}}(a_i)$ denote the second smallest component of $(a_1, \ldots, a_n)$, and denote $I = \{1, \ldots, n\}$ then

$$\sum_{j \in \mathcal{T}} u_j^*(x)(f_j(x) + \text{div } p_j^*(x)) + \sum_{i \in I \setminus \mathcal{T}} u_i^*(f_i(x)$$
$$+ \text{div } p_i^*(x))$$
$$= \sum_{j \in \mathcal{T}} u_j^*(x) \min_{i \in I}(f_i(x) + \text{div } p_i^*(x))$$
$$+ \sum_{i \in I \setminus \mathcal{T}} u_i^*(x)(f_i(x) \, \text{div } p_i^*(x))$$
$$\geq \sum_{j \in \mathcal{T}} u_j^*(x) \min_{i \in I}(f_i(x) + \text{div } p_i^*(x))$$
$$+ \sum_{j \in I \setminus \mathcal{T}} u_j^*(x) \min_{i \in I}^{2\text{nd}}(f_i(x) + \text{div } p_i^*(x)),$$
$$= (1 - \varepsilon) \min_{i \in I}(f_i(x) + \text{div } p_i^*(x))$$
$$+ \epsilon \min_{i \in I}^{2\text{nd}}(f_i(x) + \text{div } p_i^*(x))$$
$$= \min_{i \in I}(f_i(x) + \text{div } p_i^*(x))$$
$$+ \varepsilon\big(\min_{i \in I}^{2\text{nd}}(f_i(x) + \text{div } p_i^*(x))$$
$$- \min_{i \in I}(f_i(x) + \text{div } p_i^*(x))\big)$$
$$> \min_{i \in I}(f_i(x) + \text{div } p_i^*(x)).$$

Therefore, integrating over all $x \in \Omega$

$$E(u^*, p^*) = \sum_{i=1}^{n} \int_{\Omega} u_i^*(x)(f_i(x) + \text{div } p_i^*(x))dx$$
$$> \int_{\Omega} \min_{i \in I}(f_i(x) + \text{div } p_i^*(x))dx = E(p^*),$$

a contradiction to the fact that (u^*, p^*) is an optimal primal-dual pair of (13). □

Then it follows directly, from Theorem 1, that

Proposition 1 *Let p^* be one optimum of the dual problem (16). If the values $(f_1(x) + \text{div } p^*(x), \ldots, f_n(x) +$*

div $p^*(x)$) *have a unique minimum at all x in* Ω, *then the primal variable* u^* *given by* (19) *is a binary global optimum to the primal problem* (5) *and to the original non-convex Potts model* (4).

Proof By Prop (1) any primal-dual pair must satisfy (20) and (21). Since the primal variable u^* given by (19) is the only variable that satisfies these constraints for the given p^*, it follows that (u^*, p^*) is an optimal primal-dual pair by the existence of such a pair. Hence u^* is an optimum of the primal problem (5).

That u^* is also a global optimum of (4) 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.   □

In the case where there are two minimums of ($f_1(x) +$ div $p_1^*(x), \ldots, f_n(x) +$ div $p_n^*(x)$) at some positions $x \in \Omega$, it is also possible to prove existence of a global optimum

Proposition 2 *Let* p^* *be one optimum of the dual problem* (16). *If the values* ($f_1(x) +$ div $p^*(x), \ldots, f_n(x) +$ div $p^*(x)$) *have at most two minimums at all x in* Ω, *then there exists binary primal variables* u^* *which are globally optimal to the primal problem* (5) *and the original nonconvex problem* (4).

The proof of Proposition 2 is given in the appendix. 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 Nikolova et al. (2006) and Strang (1983), who gave a different proof of existence of binary minimizers for relaxed two phase problems.

Further analysis in the case the number of minimums is greater than 2 at some positions $x \in \Omega$ will be subject of future research.

In order to deal with non-differentiability of the dual energy functional (16), a smooth approximation functional is introduced in the next section. This smooth model tends to favor solutions where at each $x \in \Omega$, ($f_1(x) +$ div $p_1^*(x), \ldots, f_n(x) +$ div $p_n^*(x)$) has a unique minimum. Therefore, a binary primal solution can be recovered by the scheme (19). Since there is an approximation between the smooth and non-smooth models, one cannot guarantee the obtained solutions are exact. Experiments show that such solution may better (have lower energy) than simply solving the relaxed primal problem exactly and then threshold the result like (7).

3 Smoothed Models

In order to solve the optimal labeling problem (5) through its nonsmooth dual model (16), we propose a smoothing method in this section, which leads to the *smoothed primal-dual model* and *smoothed dual model*, associated with (13) and (16). 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_∞ of a proper convex function $g(u)$ is also a proper convex function, positively homogeneous and defined in an approximation way (Rockafellar 1970; Teboulle 2007) as

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

For example,

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

and

$$g(u) = \log \sum_{j=1}^{k} e^{u_j}, \qquad g_\infty(z) = \max_{1 \le j \le k} z_j. \qquad (22)$$

We use an example to show the smoothing effects of the Log-Sum exponential function (22) for the highly nonsmooth function $\max_{1 \le j \le k} z_j$. In Fig. 2, the nonsmooth function $f(x) = \max(1 - x, x)$ is given in the first graph. We use the Log-Sum exponential function (22) 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 (22) to approximate the min function in (16) by choosing a small parameter $s > 0$. In this way, the nonsmooth optimization problem (16) can be approximated by

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

$$(23)$$

We call the new optimization problem (23) the *smoothed dual model* in comparison to the original dual one (16).

Fig. 2 The *left graph*, the *black and bold line*, gives the function $f(x) = \max(1-x, x)$, in the interval $x \in [0, 1]$. 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*
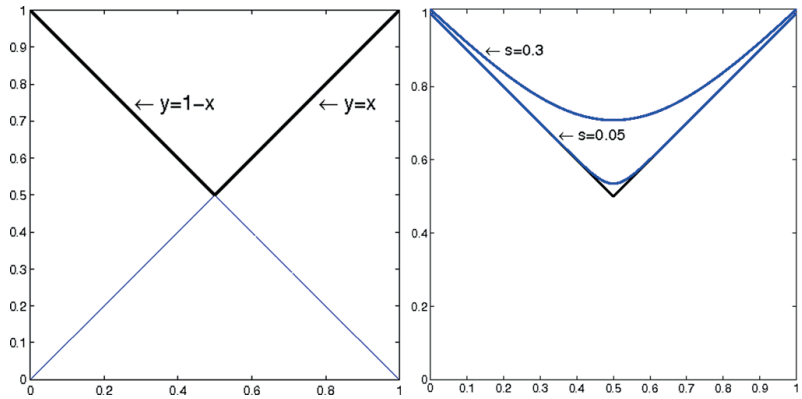
3.2 Equivalent Smoothed Models and Maximum Entropy Labelings

This section derives primal-dual and primal formulations of the smoothed dual model and reveals an interesting relationship to maximum entropy clustering algorithms. Actually, it is well known that the smooth log-sum function has an identical expression (Rockafellar 1970):

Lemma 1 *For any given $\mu \in \triangle_+$ and $h \in \mathbb{R}^n$,*

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

Let $\mu_i = 1/n$, $i = 1 \ldots n$. By the results of Lemma 1, we see that the *smoothed dual model* (23) is just equivalent to

$$\max_{p_i \in C_\alpha} \min_{u \in S} E_s(u, p)$$
$$= \int_\Omega \left\{ \sum_{i=1}^{n} u_i (f_i + \operatorname{div} p_i) + s \sum_{i=1}^{n} u_i \log u_i \right\} dx. \quad (24)$$

In view of the primal-dual model (13), the energy functional in the optimization problem (24) is just the energy functional of (13) plus an entropy-penalizing term. Such entropy penalization provides a proper regularization or smoothing of the original function. We, likewise, call the optimization problem (24) the *smoothed primal-dual model*.

Correspondingly, optimizing the dual variables p in (24) leads to the equivalent *smoothed primal model*:

$$\min_{u \in S} E_s^P(u) = \int_\Omega \left\{ \sum_{i=1}^{n} u_i f_i + \alpha |\nabla u_i| + s u_i \log u_i \right\} dx. \quad (25)$$

Clearly, the positive value s here works as a penalization parameter. When s approaches 0, the optimization problem (24) approaches the original primal-dual problem (13)

and the smoothed primal model (25) approaches the non-smooth version (5). At this, the smoothed primal-dual model (24) shares the same formulation of the *maximum entropy clustering algorithms* (Rose 1998). To this end, we also call our smoothing approach given by (23) or (24) the method of *maximum entropy labelings*.

3.3 Approximation Bounds of Smoothed Models

Some error bounds of the smoothed models are derived in this section. In fact, the Log-Sum exponential function gives the following approximation bound on the maximum function $\max_{1 \le i \le k} z_i$ (Teboulle 2007).

Lemma 2 *For each $\mu \in \triangle_+$, the following inequalities hold,*

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

Moreover, for $s > 0$

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

The proof is referred to (Teboulle 2007).

Then in view of Lemma 2, we have the approximation bound of the smoothed dual model:

Proposition 3 *For any $s > 0$, the smoothed dual model (23) gives an approximation of (16), which has the bound:*

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

where the functions $E^D(p)$ and $E_{s>0}^D(p)$ are the energy functional of (16) and (23) respectively, $|\Omega|$ is the area of the domain Ω.

Proof Define the function $G_s(x)$ as

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

i.e. the component function of (23) to be integrated.

Let $\mu_i = 1/n$ and $z_i = -(f_i(x) + \text{div } p_i(x))$ for each $x \in \Omega$. By Lemma 2, we have

$$\sum_{i=1}^{n} \frac{f_i(x) + \text{div } p_i(x)}{n} \geq s \log n + G_s(x)$$

$$\geq \min_{1 \leq i \leq k} (f_i(x) + \text{div } p_i(x)).$$

Therefore,

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

and

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

On the other hand, through Lemma 1 and (24), we have

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

Hence

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

then

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

By Proposition 3, the approximation bound of the smoothed model (23) depends on the smoothing parameter s. Hence by choosing s small enough, the smoothed dual model (23) solves the original nonsmooth dual model (16) within an expected error bound.

4 Algorithm

The smooth energy function considered in the smoothed dual model (23) allows to build up an efficient and simple numerical scheme over the dual variables $p_i(x), i = 1 \ldots n$. In order to maximize the energy functional (23), we propose a projected gradient algorithm, see Algorithm 1, which contains the same steps as the algorithms suggested in (Chambolle 2004; Kiwiel 1995).

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

Algorithm 1 Projection-based Algorithm

– Let $\delta > 0$ be chosen as some suitable step-size and let $p_i^0, i = 1, \ldots, n$ be chosen as any starting values (e.g. all identically 0), set $k = 0$ then start;
  – Compute

$$u_i^k = \frac{e^{\frac{-f_i - \text{div } p_i^k}{s}}}{\sum_{i=1}^{n} e^{\frac{-f_i - \text{div } p_i^k}{s}}}, \quad i = 1, \ldots, n; \qquad (26)$$

  – Update $p_i^{k+1}, i = 1, \ldots, n$ by

$$p_i^{k+1} = \text{Proj}_{C_\alpha}(p_i^k + \delta \nabla u_i^k), \quad i = 1, \ldots, n,$$

  where Proj_{C_α} is the projection operator to the convex set C_α;
– Let $k = k + 1$ and go to iteration $k + 1$ until convergence.
– When convergence is achieved, the primal variable u is recovered by

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

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

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

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

$$p_i^{k+1} = \text{Proj}_{C_\alpha}(p_i^k + \delta \nabla u_i^k), \quad i = 1, \ldots, n.$$

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

– **Gradient-Descent Step**, compute

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

  where ∇u_i^k is the gradient of the energy functional of (23)
– **Projection Step**, compute the projection to the convex set C_α:

$$p_i^{k+1} = \text{Proj}_{C_\alpha}(\tilde{p}_i^{k+1}), \quad i = 1, \ldots, n.$$

This algorithm can also be seen as a forward-backward splitting algorithm. Convergence proofs for such algorithms have been established in (Combettes and Wajs 2005).

4.1 Application to Image Segmentation

An important application of Potts model is image segmentation into several regions. Our numerical experiments will mainly focus on the application to image segmentation to validate the new optimization method, since the results are easily visualized and evaluated. Let $I(x)$ be the image intensity of the input image at $x \in \Omega$. One possibility for the data term is the Mumford-Shah/Chan-Vese functional. Given n mean intensities $c_i \in \mathbb{R}$, $i = 1, \ldots, n$, a data term can then be constructed as

$$f_i(x) = |I(x) - c_i|^\theta, \quad \forall x \in \Omega. \tag{27}$$

The complete algorithm for segmenting Ω into n regions Ω_i, $i = 1, \ldots, n$ is given in Algorithm 2. The mean intensities c_i, $i = 1, \ldots, n$ are of course unknown in advance, but can be estimated by some heuristic, e.g. the isodata algorithm (Velasco 1980). The image domain Ω and differential operators div and ∇ are assumed to be discrete in practice. In our experiments we apply the mimetic finite difference method, the details are provided in the next section.

Algorithm 2 Image segmentation into n regions

– Let I be the input image, and estimate the mean intensities c_i, $i = 1, \ldots, n$ (e.g. by isodata algorithm (Velasco 1980))
– For each $x \in \Omega$ compute $f_i(x) = |I(x) - c_i|^\beta$.
– Run Algorithm 1 until convergence
– For $i = 1, \ldots, n$ obtain region Ω_i by

$$\Omega_i = \{x \in \Omega | u_i(x) = 1\}$$

4.2 Implementation

In this work, we apply the *mimetic finite-difference method* (Hyman and Shashkov 1997a, 1997b) to build up the relevant discretization scheme. The scheme is described in 2-D for simplicity, but can also straight forwardly be generalized to N-D. 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. 3:

– 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. 3).
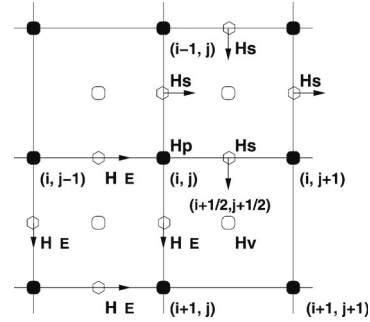

Fig. 3 Definitions of finite-dimensional spaces of scalar fields, H_V and H_P, and vector fields, H_E and H_S, on a 2-D square grid

– 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. 3);
– 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. 3);
– 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. 3).

We will implement our numerical scheme mainly by applying the H_V space for 2-D scalar fields and the H_S space for vector fields. By the *mimetic finite-difference method*, the inner product between two vectors $u, v \in H_V$ and $p, q \in H_S$ are defined by

$$\langle u, v \rangle_{H_V} := \sum_{(\alpha, \beta)} u_{\alpha, \beta} v_{\alpha, \beta};$$

and

$$\langle p, q \rangle_{H_S} := \sum_{(\alpha, \beta)} \frac{1}{2}(p^b q^b + p^r p^r + p^t q^t + p^l q^l)_{(\alpha, \beta)},$$

where (α, β) denotes one cell in the grid and (p^b, p^r, p^t, p^l) are the values of the vector field p at four sides of the cell (α, β): bottom side, right side, top side and left side respectively.

For $u \in H_V$ and $p \in H_S$, the corresponding divergence operator $\mathbb{D}iv^*$ is the dual first-order differential operator which maps any vector field in H_S to a scalar field in H_V, i.e. $\mathbb{D}iv^* : H_S \to H_V$. It is simply defined by

$$(\mathbb{D}iv^* p)_{(\alpha, \beta)} := p^b + p^r - p^t - p^l.$$

The gradient operator $\mathbb{G}^*$ is also a dual first-order differential operator which maps any scalar field in H_V to a vector

field in H_S, i.e. $\mathbb{G}^* : H_V \to H_S$. It is defined as

$$(\mathbb{G}^* u)^t_{(\alpha,\beta)} := u_{(\alpha,\beta)} - u_{(\alpha,\beta-1)},$$

$$(\mathbb{G}^* u)^b_{(\alpha,\beta)} := u_{(\alpha,\beta+1)} - u_{(\alpha,\beta)},$$

$$(\mathbb{G}^* u)^l_{(\alpha,\beta)} := u_{(\alpha,\beta)} - u_{(\alpha-1,\beta)},$$

$$(\mathbb{G}^* u)^r_{(\alpha,\beta)} := u_{(\alpha+1,\beta)} - u_{(\alpha,\beta)}.$$

Now we will explain how to implement the two steps of the algorithm: gradient step and projection step.

4.2.1 Gradient Step

At each iteration of the algorithm, given $p_i \in H_S$, $f_i \in H_V$, $i = 1, \ldots, n$, we have

$$\omega_i(\alpha,\beta) = \exp \frac{-f_i(\alpha,\beta) - (\mathbb{D}iv^* p_i)_{(\alpha,\beta)}}{s}, \quad i = 1, \ldots, n,$$

then

$$\rho_i(\alpha,\beta) = \frac{\omega_i(\alpha,\beta)}{\sum_{i=1}^n \omega_i(\alpha,\beta)}, \quad i = 1, \ldots, n,$$

for each cell (α, β).

Therefore, the gradient of ρ_i $i = 1, \ldots, n$ is given by

$$d_i = \mathbb{G}^* \rho_i, \quad i = 1, \ldots, n, \tag{28}$$

and

$$\tilde{p}_i = p_i + \delta d_i, \quad i = 1, \ldots, n.$$

4.3 Projection Step

Recall the convex constraint set of the dual variables (11). In the discrete setting, by the *mimetic finite-difference method*, any vector field $p \in C_\lambda$, at each cell (α, β) should satisfy

$$\ell_{\alpha,\beta}(p) := \sqrt{\frac{1}{2}\big((p^b)^2 + (p^r)^2 + (p^t)^2 + (p^l)^2\big)_{(\alpha,\beta)}} \leq \lambda.$$

The projection of any vector field p to the convex set C_λ can be approximated by the following two steps:

– Define $\tilde{\ell}(p) \in H_V$:

$$\tilde{\ell}_{\alpha,\beta}(p) := \begin{cases} \lambda/\ell_{\alpha,\beta}(p) & \text{when } \ell_{\alpha,\beta}(p) \geq \lambda, \\ 1 & \text{when } \ell_{\alpha,\beta}(p) < \lambda, \end{cases}$$

and define the discrete vector field $q \in H_S$:

$$q^t_{(\alpha,\beta)} := (\tilde{\ell}_{(\alpha,\beta)} + \tilde{\ell}_{(\alpha,\beta-1)})/2,$$

$$q^b_{(\alpha,\beta)} := (\tilde{\ell}_{(\alpha,\beta+1)} + \tilde{\ell}_{(\alpha,\beta)})/2,$$

$$q^l_{(\alpha,\beta)} := (\tilde{\ell}_{(\alpha,\beta)} + \tilde{\ell}_{(\alpha-1,\beta)})/2,$$

$$q^r_{(\alpha,\beta)} := (\tilde{\ell}_{(\alpha+1,\beta)} + \tilde{\ell}_{(\alpha,\beta)})/2;$$

– $\tilde{p} := \text{Proj}_{C_\lambda}(p) \in H_S$ is computed by:

$$\tilde{p}^t_{(\alpha,\beta)} := p^t_{(\alpha,\beta)} q^t_{(\alpha,\beta)}, \qquad \tilde{p}^b_{(\alpha,\beta)} := p^b_{(\alpha,\beta)} q^b_{(\alpha,\beta)},$$

$$\tilde{p}^l_{(\alpha,\beta)} := p^l_{(\alpha,\beta)} q^l_{(\alpha,\beta)}, \qquad \tilde{p}^r_{(\alpha,\beta)} := p^r_{(\alpha,\beta)} q^r_{(\alpha,\beta)}.$$

4.3.1 Step to Compute u

When the algorithm converges to some optimal p_i^* $i = 1, \ldots, n$, evaluate u_l by

$$u_l(\alpha, \beta) = \begin{cases} 1 & \text{if } l = \arg\min_{i=1,\ldots,n}(f_i + \mathbb{D}iv^* p_i)_{(\alpha,\beta)}, \\ 0 & \text{otherwise}. \end{cases}$$

5 Numerical Experiments

We demonstrate the performance of the smoothed dual model by several experiments and compare with established methods. Alpha expansion and alpha-beta swap (Boykov et al. 2001) 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 Euclidean length of the boundaries. Because of this difference, energy comparison is not straight forward. However, there exists a result which allows to approximate the Euclidean curve length on a discrete grid. This result is called the Cauchy-Crofton formula and was specialized for computer vision problems in Boykov and Kolmogorov (2003). In short, it gives a formula for edge weights between neighboring grid points such that the discrete boundary length converges to the Euclidean 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, i.e. it can be used to compare energy of the thresholded solutions. Secondly, we evaluate quality and efficiency with the approaches of (Zach et al. 2008; Lellmann et al. 2009). Energy plots for all experiments can be found in Fig. 10. The final energies of the different methods are plotted as a function of the regularization parameter α. Some comparisons are also made to the very recent convex relaxation approach (Pock et al. 2008b) 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. The relaxation (Pock et al. 2008b) can be shown to be tighter, but is more computationally complex, especially when the number of labels is large.

In experiments where the correct solution is known, we have also compared the percentage of misclassified pixels,

Table 1 Percentage of misclassified pixels for experiment 1–4 (α-expansion and $\alpha - \beta$-swap implemented with 4 neighborhood system)

	α-exp	$\alpha - \beta$-swap	Lellmann et al.	Dual
Experiment 1	8.89	6.12	–	5.51
Experiment 2	1.17	1.17	–	1.06
Experiment 3	7.42	15.72	12.30	11.72
Experiment 4	6.64	7.23	6.25	5.86

Table 1. The regularization parameter α 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++ (Boykov et al. 2001). The input images in Figs. 4 and 9 was first used by Pock et al. (2008a), and the input images in Figs. 7 and 8 was first used by Lellmann et al. (2009).

5.1 Qualitative Evaluation

Figure 4 presents the full experiments on the flower image 1 from the introduction. 10 phases/labels have been used, with color data fidelity

$$f_i = \sum_{j=1}^{3} |I - c_i^j|, \quad i = 1, \ldots, 10, \tag{29}$$

where $\{\mathbf{c}_i\}_{i=1}^{10}$ are predefined color vectors. In Fig. 4(b)–(g) a low regularization ($\alpha = 10$) has been chosen. In Fig. 4(f)–(e), a higher regularization ($\alpha = 40$) is used. Alpha expansion and alpha beta swap leads to metrication errors, which is particularly visible with 4 neighbors and low regularization (b) and 8 neighbors and high regularization (f). In addition artifacts are introduced as the energy is not minimized exactly, see e.g. the transition between flower and sky. In terms of energy, the smoothed dual model outperforms the graph cut based approaches, see Fig. 10(a), especially when α is large. The results in the introduction, Fig. 1, were generated with the largest α in the energy plot. In Fig. 4(c) a comparison with the recent method of Pock et al. is made. Their method seems to recover almost integer valued solutions up to some blurring of the boundaries.

Some artificial examples are presented next in experiment 1–4, Figs. 5–8. The leftmost gray scale image I is to be classified into 4 classes by using the L^1 norm in the data fidelity term

$$f_i = |I - c_i|, \quad i = 1, \ldots, 4, \tag{30}$$

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 implemented
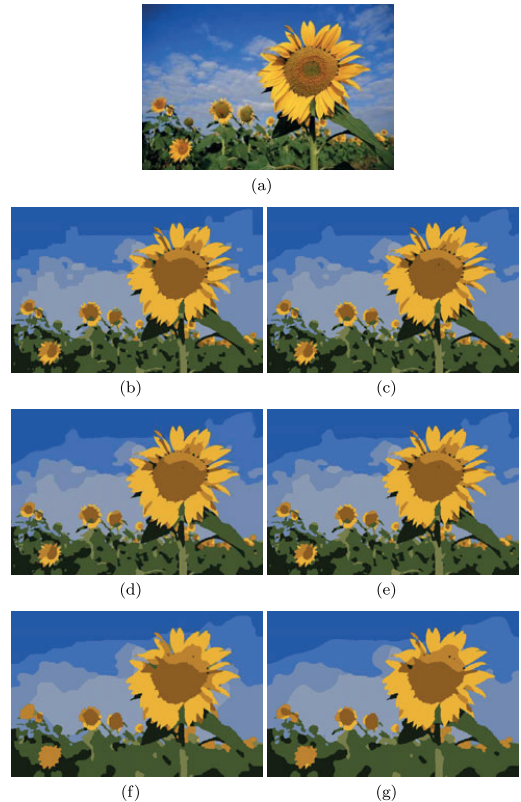

Fig. 4 (**a**) Input, (**b**) alpha expansion 4 neighbors, (**c**) alpha expansion 8 neighbors, (**d**) Pock et al. (**e**) dual model. (**f**)–(**g**): $\alpha = 40$, (**f**) alpha expansion 8 neighbors, (**g**) dual model
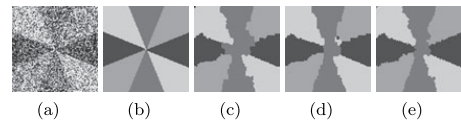
Fig. 5 Experiment 1: (**a**) Input, (**b**) ground truth, (**c**) alpha expansion, (**d**) alpha-beta swap, (**e**) dual model. Size 100×100
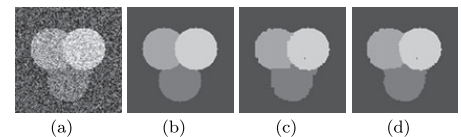
Fig. 6 Experiment 2: (**a**) Input, (**b**) ground truth, (**c**) alpha expansion, (**d**) dual model. Size: 100×100
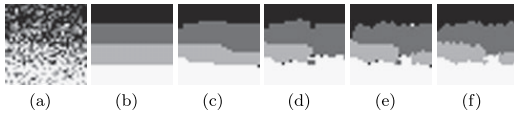
Fig. 7 Experiment 3: (**a**) Input, (**b**) ground truth, (**c**) alpha expansion, (**d**) alpha-beta swap, (**e**) Lellmann et al., (**f**) dual model. Size: 32×32
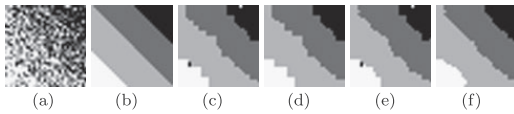
Fig. 8 Experiment 4: (**a**) Input, (**b**) ground truth, (**c**) alpha expansion, (**d**) alpha-beta swap, (**e**) Lellmann et al., (**f**) dual model. Size: 32×32
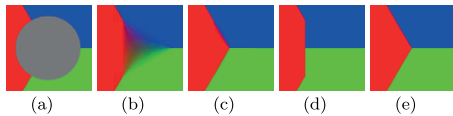

Fig. 9 (**a**) Input, (**b**) Lellmann et al., (**c**) Pock et al., (**d**) Alpha expansion (**e**) dual model

with 4 neighbors, 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 4-neighborhood 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 performs best. For energy plots, see Fig. 10, where we also have used 8 neighborhoods in the discrete models. 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 the approaches (Zach et al. 2008; Lellmann et al. 2009). This is particularly visible in Fig. 10(c).

The advantage of the smoothing is illustrated in the next example, Fig. 9, 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 (29) 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. In this example we expect that for the non-smooth model $(f_1(x) + \operatorname{div} p_1^*(x), \ldots, f_3(x) + \operatorname{div} p_3^*))$ 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 final result is shown in Fig. 9, where we also compare with other methods. The difference

between the components of $(f_1(x) + \operatorname{div} p_1^s(x), \ldots, f_3(x) + \operatorname{div} p_3^s))$ is small near the center of the image, hence it is difficult to verify whether the reconstructed solution is also globally optimal to the original problem, although it can be verified visually, since the global solution is known. As seen in the following subfigures, the approach of Lellmann et al. does not recover a binary solution. Alpha expansion yields a binary, but incorrect result, Fig. 9(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. (2008a), who first tested their method on this image. As can be seen, they can almost recover the integer valued global minimum, up to some blurring of the boundaries. After thresholding, they are also able reconstruct the triple junction. Numerical calculations for triple junctions have also been tested in Li and Tai (2007) showing that the piecewise constant level set method is able to produce 120 degrees for the junctions.

Figure 12(b) shows the result of 4 class segmentation of a brain MRI image. One would like to classify the input image in Fig. 12(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

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

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 (Bae and Tai 2009a). This algorithm finds a local minimum with respect the constant values. For energy plots, see Fig. 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 Fig. 10: we can obtain binary solutions of lower energy than the approaches of Zach et al. (2008), Lellmann et al. (2009).

5.2 Evaluation of Efficiency and Convergence

We will now compare the cpu time and convergence with the approaches (Zach et al. 2008; Lellmann et al. 2009). In order to deal with the simplex constraint in the primal optimization problem (5), an alternating optimization approach was used in (Zach et al. 2008) where one solve for $k = 1, \ldots$

$$u^{k+1} = \arg\min_u E_1(u)$$
$$= \sum_{i=1}^n \int_\Omega \frac{1}{2\tau}(u_i - v_i^k)^2 + |\nabla u_i| \, dx, \tag{31}$$

Fig. 10 Energy plot as a function of regularization parameter α of binary solutions obtained by each method. *Red*: smoothed dual model, black dotted: alpha expansion 8 neighbors, light blue: alpha expansion 4 neighbors, green +: alpha-beta swap 8 neighbors, blue *x*: Zach et al. (2008). (**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. (**a**) Is a typically difficult example with a large number of labels, where the smoothed dual model clearly performs best
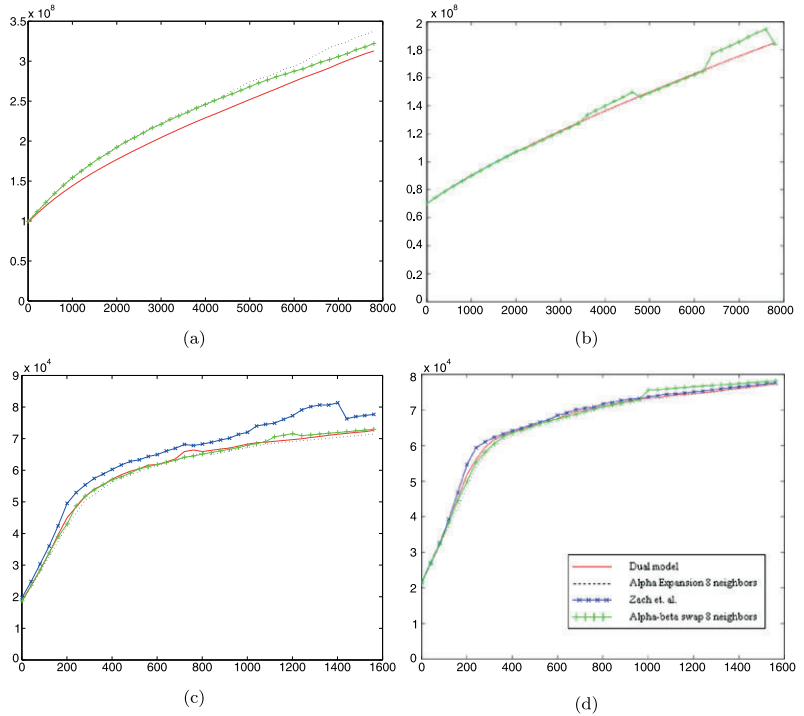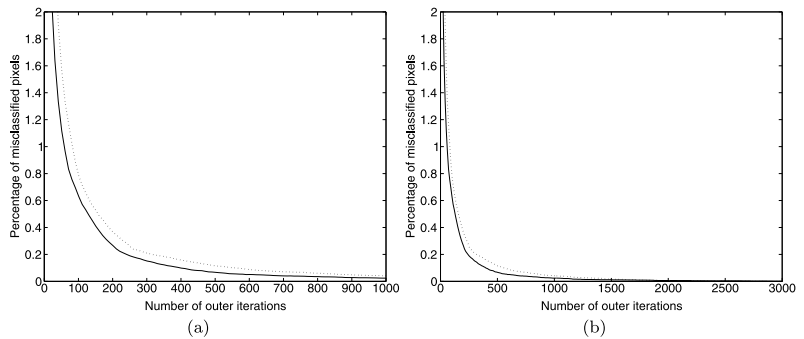

Fig. 11 Convergence rate for flower image. *Solid*: percentage of misclassified pixels as a function of the iteration count for the smoothed dual model. *Dotted*: percentage of misclassified pixels as a function of the *outer* iteration count for the Douglas-Rachford splitting approach

$$v^{k+1} = \arg\min_{v \in S} E_2(v)$$

$$= \sum_{i=1}^{n} \int_{\Omega} \frac{1}{2\tau} (u_i^{k+1} - v_i)^2 + \alpha f_i v_i \, dx, \qquad (32)$$

Here τ is some small parameter. The second problems can be optimized pointwise and has closed form solutions. However, the first subproblems are TV optimization problems, and must be solved by some iterative technique such as Chambolle's algorithm (Chambolle 2004). In Lellmann et

al. (2009) a Douglas-Rachford splitting scheme was applied to deal with the simplex constraint. After some derivations, such a scheme can be written

$$u^k = \arg\min_u \frac{1}{2} \sum_{i=1}^{n} \int_{\Omega} (u_i - (z_i^k - \tau f_i))^2 dx,$$

$$+ (\tau\alpha) \int_{\Omega} |\nabla u_i| dx \qquad (33)$$

Table 2 Number of iterations to reach exact solution

	Zach et al.	Douglas Rachford	Dual
Experiment 3	1178 ($\times$30)	610 ($\times$30)	481
Experiment 4	1504 ($\times$30)	540 ($\times$30)	425

$$w^k = \arg\min_{w \in S} \frac{1}{2\tau} \int_\Omega (w - (2u^k - z^k))^2 dx,$$

$$z^{k+1} = z^k + w^k - u^k. \tag{34}$$

As we see, this scheme also involves a substep (33) where n TV minimization problems need to be solved iteratively. The second subproblems (34) have closed form solutions and can be solved efficiently. In contrast to (31), (32), such a scheme can be proved to convergence provided the subproblems are solved exactly. As we see, 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 (Zach et al. 2008; Lellmann et al. 2009). When τ is low, the problem is solved with high accuracy, but more iterations are required. Therefore, one could say τ 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 δ 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. The approaches (Lellmann et al. 2009; Zach et al. 2008) require more parameters, like the outer time step τ, inner time step, accuracy of solving inner problem etc. We have done our best to optimize these parameters such that the algorithms converge as fast as possible.

Convergence is measured as the number of iterations required to reach the exact thresholded solutions. Since there are finitely many possibilities for such thresholded solutions, converge will occur in a finite number of steps. The exact solutions are determined by running each method for 5000 iterations. The iteration counts for the experiments in Figs. 7 and 8 are shown in Table 2. Around 400 iterations are required for the smoothed dual model. The Douglas-Rachford method requires slightly more outer iterations, but also contains an inner loop for each such outer iteration. As indicated in the parenthesis, the inner problems are solved approximately by 30 iterations of Chambolle's algorithm. Overall our approach is therefore significantly faster. The splitting approach (31), (32) falls even further behind. The flower image in Fig. 11 is larger (508×336), hence such a pixel-wise termination criterion is more strict. Figure 11 shows the percentage of incorrect pixels compared to the exact solution, as a function of the iteration count for the dual model (solid) and the outer iteration count of the Douglas-Rachford splitting approach (dotted). In this example the dual model also
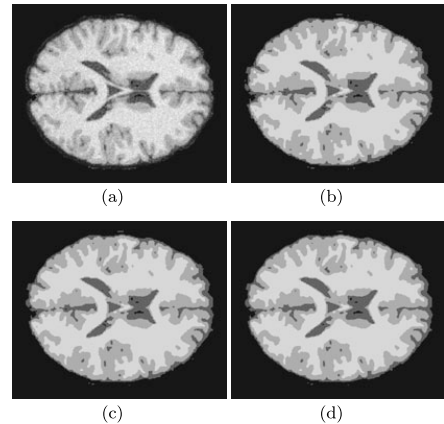
Fig. 12 (**a**) Input, (**b**) alpha expansion 4 neighbors, (**c**) alpha expansion 8 neighbors, (**d**) dual model. Size: 709×591

outperforms the Douglas Rachford scheme in terms of outer iterations. After 70 iterations, more than 99% of the pixels have reached steady state. After 300 iterations more than 99.8% have reached steady state.

We should mention that some other work has appeared recently, after this paper was submitted and made public as a preprint, on optimization of the problem 5. In Yuan et al. (2010), the dual model was interpreted as a continuous maximum flow problem. Another algorithm was derived which was shown to converge faster than the approaches of Lellmann et al. (2009), Zach et al. (2008). Another variant of the Douglas-Rachford algorithm was derived in Lellmann et al. (2010), related to the Split Bregman method (Goldstein et al. 2009), where the subproblems are instead Laplace equations, which are easier to solve than TV minimization problems.

Comparison with graph cut based alpha expansion and alpha beta swap is more difficult. We have used the highly optimized c++ implementation (Boykov et al. 2001; Boykov and Kolmogorov 2001) of these methods, while the algorithm for the dual model is implemented in a simple matlab program. There has recently been much effort on comparing continuous and discrete (graph cut) techniques in computer vision, see e.g. (Klodt et al. 2008) for an extensive discussion about two phase partitioning problems. Continuous convex optimization techniques consist mainly of floating point matrix/vector arithmetic, which is highly suited for massive parallel implementation on GPU. In contrast, combinatorial max-flow algorithms have a much more serial nature. In this regard, continuous convex optimization has been shown to be faster (Klodt et al. 2008). This is especially evident in 3D. The development of processor technology is expected to be largely of the parallel aspect in the future. Hence we see our work as more suitable for current

and future generations of hardware. Unfortunately, we don't have the resources to implement our method in such a parallel manner on GPU. 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 simple matlab implementation are as follows: For the 709×591 brain image convergence was reached in 1 minutes and 32 seconds for our implementation. For the 32×32 images in Figs. 7 and 8 convergence averaged around 2.5 seconds. For the 100×100 images in Figs. 5 and 6 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.

6 Conclusions

This paper proposed a novel duality-based approach for continuous multi-labeling problems based on a convex relaxation of Potts model. The dual model could be used to give insight into the exactness of the relaxation. Sufficient conditions were derived for when optimal solutions to the Potts model could be obtained from a dual solution to the relaxed model. Close connections between optimal labelings and geometrical clustering of spatial points were also revealed. We then suggested a smoothing method based on the log-sum exponential function, so as to deal with the nonsmooth dual problem, and indicated that such a smoothing approach leads to a novel smoothed primal-dual model and suggests labelings with maximum entropy. A new expectation maximization like algorithm was proposed based on smoothed dual model which was shown to be superior in efficiency compared to earlier approaches. Numerical experiments also showed that our approach could outperform several competitive approaches in various aspects, such as lower energies and better visual quality.

Appendix: Proof of Proposition 2

Before we prove Proposition 2, we give the following result

Proposition 4 *Given a bounded scalar field u defined on Ω, we assume, without loss of generality, $0 \le u(x) \le 1$ for all $x \in \Omega$. If a vector field p^* maximizes the integral $\int_\Omega u \, \mathrm{div} \, p \, dx$ over the convex set*

$$C_\alpha := \{p \,|\, |p(x)| \le \alpha, \ p_n = 0\}, \tag{35}$$

then for almost every level set u^γ of u

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

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

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

$$\int_\Omega |\nabla u| dx = \int_\Gamma \int_\Omega |\nabla u^\gamma| dx d\gamma. \tag{36}$$

By applying this formula we can deduce

$$\int_\Omega u \, \mathrm{div} \, p^* dx = \int_\Omega |\nabla u| dx = \int_\Gamma \int_\Omega |\nabla u^\gamma| dx d\gamma$$

$$= \int_\Gamma \left(\max_{p \in C_\alpha} \int_\Omega u^\gamma \, \mathrm{div} \, p \, dx \right) d\gamma. \tag{37}$$

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

$$\int_\Omega u \, \mathrm{div} \, p^* dx = \int_\Omega \left(\int_\Gamma u^\gamma(x) d\gamma \right) \mathrm{div} \, p^*(x) dx$$

$$= \int_\Gamma \int_\Omega u^\gamma \, \mathrm{div} \, p^* dx d\gamma. \tag{38}$$

Therefore, combining (37) and (38)

$$\int_\Gamma \int_\Omega u^\gamma \, \mathrm{div} \, p^* dx d\gamma = \int_\Gamma \left(\max_{p \in C_\alpha} \int_\Omega u^\gamma \, \mathrm{div} \, p \, dx \right) d\gamma. \tag{39}$$

This equality (39) together with the fact that for any $\gamma \in [0, 1]$

$$\int_\Omega u^\gamma \, \mathrm{div} \, p^* dx \le \max_{p \in C_\alpha} \int_\Omega u^\gamma \, \mathrm{div} \, p \, dx \tag{40}$$

implies that

$$\int_\Omega u^\gamma \, \mathrm{div} \, p^* dx = \max_{p \in C_\alpha} \int_\Omega u^\gamma \, \mathrm{div} \, p \, dx$$

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

$$L^\gamma = \int_\Omega |\nabla u^\gamma| dx = \max_{p \in C_\alpha} \int_\Omega u^\gamma \, \mathrm{div} \, p \, dx. \qquad \square$$

Now we give the proof of Proposition 2.

Proof Denote by $\Omega_{k,j}$ the set of points $x \in \Omega$ where $k, j = \arg \min_i (f_i(x) + \mathrm{div} \, p_i^*(x))$. Then $\bigcup_{k,j=1, k \ne j}^n \Omega_{k,j}$ is the set of points x where $(f_i(x) + \mathrm{div} \, p_i^*(x))$ has two minimums.

Let u^d be an optimal primal variable such that (u^d, p^*) is an optimal primal-dual pair. Consider one such non-empty subdomain $\Omega_{k,j} \neq \emptyset$. By Theorem 1

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

and

$$E^D(p^*) = E(u^d, p^*) = E^P(u^d). \tag{41}$$

Obviously, p_i^* maximizes the integral

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

over the convex set C_α (35).

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

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

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

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

and

$$\tilde{u}_k(x) = u_k^d(x), \quad x \in \Omega \backslash \Omega_{k,j}.$$

Then $\tilde{u}_k$ is binary in $\Omega_{k,j}$. By such a configuration of $\tilde{u}_k$, let

$$\tilde{u}_j(x) = 1 - \tilde{u}_k(x), \quad \forall x \in \Omega_{k,j}$$

and

$$\tilde{u}_j(x) = u_j^d(x), \quad \forall x \in \Omega \backslash \Omega_{k,j},$$

which is also binary in $\Omega_{k,j}$.

Such a choice of $\tilde{u}_i(x)$, $i = 1, \ldots, n$ doesn't change the total energy of (16), i.e.

$$\begin{aligned} E(\tilde{u}, p^*) &= \sum_{i=1}^n \int_\Omega \tilde{u}_i(x)(f_i(x) + \operatorname{div} p_i^*(x)) dx \\ &= \int_\Omega \min \left(f_1(x) + \operatorname{div} p_1^*(x), \ldots, f_n(x) \right. \\ &\quad \left. + \operatorname{div} p_n^*(x) \right) dx \\ &= E^D(p^*). \end{aligned} \tag{42}$$

It remains to show that $E(\tilde{u}, p^*) = E(\tilde{u})$. By Proposition 4, p_k^* maximizes the integral $\int_\Omega \tilde{u}_k \operatorname{div} p_k dx$ over the convex set C_α. By the fact

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

then $\tilde{u}_j(x)$ is the $1 - \gamma$ level set of $u_j^d(x)$. Therefore p_j^* also maximizes the integral $\int_\Omega \tilde{u}_j \operatorname{div} p_j dx$ over the convex set C_α. By the above facts, the total energy (42) 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). This process can be repeated for each nonempty subdomain $\Omega_{k,j}$ to construct a binary minimum of (5). □

References

Bae, E., & Tai, X. C. (2009a). Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. In *Scale space and variational methods in computer vision* (pp. 1–13). Berlin: Springer.

Bae, E., & Tai, X.-C. (2009b). Efficient global minimization for the multiphase Chan-Vese model of image segmentation. In *Energy minimization methods in computer vision and pattern recognition (EMMCVPR)* (pp. 28–41).

Banerjee, S., Merugu, I., Dhillon, J., & Ghosh, J. (2004). Clustering with Bregman divergences. *Journal of Machine Learning Research*, 234–245.

Boykov, Y., & Kolmogorov, V. (2001). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 359–374.

Boykov, Y., & Kolmogorov, V. (2003). Computing geodesics and minimal surfaces via graph cuts. In *ICCV '03: proceedings of the ninth IEEE international conference on computer vision* (pp. 26–33). Washington: IEEE Computer Society.

Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23, 1222–1239.

Boykov, Y., Kolmogorov, V., Cremers, D., & Delong, A. (2006). An integral solution to surface evolution pdes via geo-cuts. In *ECCV* (pp. 409–422).

Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J. P., & Osher, S. (2007). Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2), 151–167.

Brown, E. S., Chan, T. F., & Bresson, X. (2009). *Convex formulation and exact global solutions for multi-phase piecewise constant mumford-shah image segmentation* (CAM-report-09-66). UCLA, Applied Mathematics, July 2009.

Brown, E. S., Chan, T. F., & Bresson, X. (2010). *A convex relaxation method for a class of vector-valued minimization problems with applications to mumford-shah segmentation* (CAM-report-10-43). UCLA, Applied Mathematics, July 2010.

Chambolle, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1), 89–97.

Chan, T. F., & Esedoglu, S. (2005). Aspects of total variation regularized l[sup 1] function approximation. *SIAM Journal of Applied Mathematics*, 65(5), 1817–1837.

Chan, T., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, 10, 266–277.

Combettes, P. L., & Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4), 1168–1200.

Ekeland, I., & Téman, R. (1999). *Convex analysis and variational problems*. Philadelphia: Society for Industrial and Applied Mathematics.

Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, *39*, 42–47.

Goldstein, T., Bresson, X., & Osher, S. (2009). *Geometric applications of the split Bregman method: segmentation and surface reconstruction* (UCLA CAM Report 09-06).

Greig, D. M., Porteous, B. T., & Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 271–279.

Hyman, J. M., & Shashkov, M. J. (1997a). Adjoint operators for the natural discretizations of the divergence, gradient and curl on logically rectangular grids. *Applied Numerical Mathematics*, *25*(4), 413–442.

Hyman, J. M., & Shashkov, M. J. (1997b). Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Computer Mathematics and Its Applications*, *33*(4), 81–104.

Ishikawa, H. (2003). Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*, 1333–1336.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: active contour models. *International Journal of Computer Vision*, *1*(4), 321–331.

Kiwiel, K. C. (1995). Proximal minimization methods with generalized Bregman functions. *SIAM Journal on Control and Optimization*, *35*, 1142–1168.

Klodt, M., Schoenemann, T., Kolev, K., Schikora, M., & Cremers, D. (2008). An experimental comparison of discrete and continuous shape optimization methods. In *ECCV '08: proceedings of the 10th European conference on computer vision* (pp. 332–345). Berlin: Springer.

Kohli, P., Pawan Kumar, M., & Torr, P. H. S. (2009). p^3 and beyond: move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(9), 1645–1656.

Kolmogorov, V. (2005). What metrics can be approximated by geocuts, or global optimization of length/area and flux. In *ICCV* (pp. 564–571).

Kolmogorov, V. (2006). Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(10), 1568–1583.

Kolmogorov, V., & Zabih, R. (2004). What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*, 65–81.

Komodakis, N., & Tziritas, G. (2007). Approximate labeling via graphcuts based on linear programming. *Pattern Analysis and Machine Intelligence*, 1436–1453.

Lellmann, J., Kappes, J., Yuan, J., Becker, F., & Schnörr, C. (2009). 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* (pp. 150–162). Berlin: Springer.

Lellmann, J., Breitenreicher, D., & Schnörr, C. (2010). Fast and exact primal-dual iterations for variational problems in computer vision. In *European conference on computer vision (ECCV)*.

Li, S. Z. (2001). *Markov random field modeling in image analysis*. New York: Springer.

Li, H., & Tai, X. C. (2007). Piecewise constant level set methods for multiphase motion. *International Journal of Numerical Analysis and Modeling*, *4*, 274–293.

Lie, J., Lysaker, M., & Tai, X.-C. (2006a). A variant of the level set method and applications to image segmentation. *Mathematics of Computation*, *75*(255), 1155–1174.

Lie, J., Lysaker, M., & Tai, X. C. (2006b). A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE Transactions on Image Processing*, *15*(5), 1171–1181.

Meyer, Y. (2001). *University lecture series: Vol. 22. Oscillating patterns in image processing and nonlinear evolution equations*. Providence: American Mathematical Society. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

Mumford, D., & Shah, J. (1989). Optimal approximation by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, *42*, 577–685.

Nikolova, M., Esedoglu, S., & Chan, T. F. (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, *66*(5), 1632–1648.

Osher, S., & Sethian, J. A. (1988). Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, *79*(1), 12–49.

Paragios, N., Chen, Y., & Faugeras, O. (2005). *Handbook of mathematical models in computer vision*. New York: Springer.

Pock, T., Schoenemann, T., Graber, G., Bischof, H., & Cremers, D. (2008a). A convex formulation of continuous multi-label problems. In *European conference on computer vision (ECCV)*, Marseille, France, October 2008.

Pock, T., Schoenemann, T., Graber, G., Bischof, H., & Cremers, D. (2008b). A convex formulation of continuous multi-label problems. In *ECCV '08: proceedings of the 10th European conference on computer vision* (pp. 792–805). Berlin: Springer.

Pock, T., Chambolle, A., Bischof, H., & Cremers, D. (2009). A convex relaxation approach for computing minimal partitions. In *IEEE conference on computer vision and pattern recognition (CVPR)*, Miami, FL.

Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophical Society*, *48*, 106–109.

Rockafellar, R. T. (1970). *Princeton mathematical series: Vol. 28. Convex analysis*. Princeton: Princeton University Press.

Rose, K. (1998). Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. In *Proceedings of the IEEE* (pp. 2210–2239).

Strang, G. (1983). Maximal flow through a domain. *Mathematical Programming*, *26*, 123–143.

Tao, W., & Tai, X. (2009). *Multiple piecewise constant active contours for image segmentation using graph cuts optimization* (UCLA cam-report 09-13).

Teboulle, M. (2007). A unified continuous optimization framework for center-based clustering methods. *Journal of Machine Learning Research*, *8*, 65–102.

Velasco, F. R. D. (1980). Thresholding using the ISODATA clustering algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, *10*(11), 771–774.

Vese, L. A., & Chan, T. F. (2002). A new multiphase level set framework for image segmentation via the mumford and shah model. *International Journal of Computer Vision*, *50*, 271–293.

Wainwright, M., Jaakkola, T., & Willsky, A. (2002). Map estimation via agreement on (hyper)trees: message-passing and linear programming approaches. *IEEE Transactions on Information Theory*, *51*, 3697–3717.

Yuan, J., Bae, E., Tai, X.-C., & Boykov, Y. (2010). A continuous maxflow approach to Potts model. In *ECCV 2010: proceedings of the 11th European conference on computer vision* (pp. 332–345). Berlin: Springer.

Zach, C., Gallup, D., Frahm, J.-M., & Niethammer, M. (2008). Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, modeling and visualization workshop (VMV)*.