A Convex Approach for Multi-phase Piecewise Constant Mumford-Shah Image Segmentation

Ethan S. Brown, Tony F. Chan, and Xavier Bresson*

Abstract

Recently, there has been significant research exploring methods to find global solutions for the piecewise constant Mumford-Shah image segmentation problem (known as the Potts model in the discrete setting). In this work, we propose a convex relaxation method to solve this non-convex problem. Our approach is based on applying a modified version of a technique due to Pock et al. [28] to a specific representation of the segmentation problem due to Lie et al. [24]. Once we have the convex optimization problem, we give an algorithm to compute a global solution. We demonstrate our algorithm on several multi-phase image segmentation problems, including a medical imaging application and a synthetic triple junction example. Even though our method cannot guarantee a global solution of the original problem, we justify the effectiveness of the approach and show the connection between our method and previous convex relaxation techniques.

1 Introduction

Image segmentation is one of the most fundamental problems in computer vision. The task is to assign a label to each point in a given image (i.e., partition the image into phases or segments) such that the labeling is optimal with respect to a particular model. It is difficult to compute a globally optimal solution because the models often lead to non-convex variational

*Authors are with the Department of Mathematics, University of California, Los Angeles. Email: {ethan,chan,xbresson}@math.ucla.edu. This work was supported in part by NSF IIS-0914580 and ONR N00014-09-1-0105.

problems. Most techniques instead devise algorithms to efficiently compute solutions which are only locally optimal.

The labels assigned during image segmentation can be defined either on a discrete grid or in a spatially continuous domain. In the discrete setting, interest for guaranteed global solutions of vision problems has grown ever since the stochastic approximate methods in the seminal work of Geman and Geman [18]. Greig et al. [19] were the first to discover that the classical Ford-Fulkerson min-cut/max-flow correspondence [17] could provide global algorithms when the set of labels is binary and the model satisfies certain criteria. The work of [20] showed that a class of multi-label problems can be globally solved, but this does not include one of the most celebrated image segmentation models: the piecewise constant Mumford-Shah model [25] (known as the Potts model [29] for the discrete problem). Recently, Bae and Tai [2] provide a global graph cut method for the Vese-Chan model, which closely approximates Mumford-Shah with a slightly different regularization term. In this paper, however, our focus is the continuous setting; see [21] for some discussion on the comparison between discrete and continuous methods in optimization problems.

For the *two-phase* case (i.e., assigning one of just two labels, foreground or background, to each point in the continuous domain), Chan et al. [12] developed an equivalent convex formulation, thereby constructing an algorithm that guarantees to find a globally optimal solution. The method relies on a relaxation of the binary labeling function which remarkably still yields a solution that is binary. However, this technique is not directly applicable to the *multi-phase* case, which has remained an open problem.

Inspired by [12], there has been significant research devoted to globally solving the multi-phase segmentation problem (or equivalent multi-labeling problems). Zach et al. [33] proposed a relaxation approach along with a decoupling to yield a convex optimization problem, but could not guarantee a global optimum of the original problem. Lellmann et al. [23] used a similar method, with a slightly different regularization and splitting method, but this had comparable limitations. Later, Pock et al. [27] used a different relaxation of the problem in a primal-dual setting. Their relaxation strictly dominates these previous approaches, in the sense that the set of functions over which their convex minimization is conducted contains that of the earlier methods. Most recently, Bae, Yuan and Tai [3] examined the dual interpretation of the relaxation of [33, 23] which provided another algorithm based on a smoothing technique.

Since the Potts model has been shown to be NP-hard in the discrete setting [5], we cannot expect to be able to compute global solutions in all cases. The goal of this paper is to provide a convex formulation and an algorithm to compute approximate global solutions for the multi-phase piecewise constant Mumford-Shah model. Our method relies primarily on two existing approaches for image segmentation and computer vision problems. First, we use the so-called *piecewise constant level set method* (PCLSM) framework of Lie et al. [24] to formulate the problem, which provides a convenient representation to enforce the constraints in the optimization. Using an augmented Lagrangian method, we are able to reduce the original minimization problem to a sequence of more tractable subproblems. Second, we use the novel *convexification* approach of Pock et al. [28] (which was motivated by the work of Ishikawa in the discrete Markov Random Field setting [20]) to tackle each subproblem. This converts the non-convex variational problem to a problem in a higher dimensional space where the objective function is convex, where a relaxation of the constraint set is made. Although the relaxation is not exact in the general case, the algorithm produces the global solution in most cases because intermediate solutions of our iterative algorithm satisfy a coarea property.

The paper is organized as follows. In Section 2, we formally introduce the segmentation problem and explain the solution of [12] for the two-phase case. In Section 3, we begin by explaining the representation we use for the multi-phase problem. We then establish a crucial augmented Lagrangian method and then invoke a convexification technique similar to [28]. Section 4 describes our algorithm, its numerical implementation, and shows some experimental results. We compare our method to previous work in this direction in Section 5, establishing a connection between our method and the convex simplex relaxation methods in [33, 23]. Finally, in Section 6, we give some concluding remarks.

2 Segmentation problem and two-phase solution

Let $I\colon \Omega \to \mathbb{R}$ be a given image with bounded image domain $\Omega \subset \mathbb{R}^d$. Mumford and Shah [25] proposed the minimization problem

$$\inf_{f,C} \int_\Omega (I(x) - f(x))^2 \, dx + \nu \int_{\Omega \setminus C} |\nabla f(x)|^2 \, dx + \mu |C|$$

to find an optimal piecewise smooth approximation of a general function (the image I), where ν and μ are fixed parameters. The edge set $C \subset \Omega$ is a closed set that defines a partition $\Omega = \cup_i \Omega_i$ such that the restrictions f_i of the function f to the segments Ω_i are differentiable. In the simplest form of the model, the function f is taken to be constant on each segment, reducing to the piecewise constant Mumford-Shah problem

$$\inf_{c_i,C} \left\{ \sum_i \int_{\Omega_i} (c_i - I(x))^2 \, dx + \mu \, |C| \right\},$$

where $f = \sum_i c_i 1_{\Omega_i}$. This is also known as the Potts model [29] in the discrete setting, originating in solid state physics, and hence we will use the names of these models interchangeably. The important property of the model is imposing a regularity measure that favors the labels of neighboring points to be identical but interprets the value of each label to be immaterial.

We will herein assume that the optimal constants c_i are known *a priori* and the number of segments (say, $k+1$) is fixed. Indeed, even making this stringent assumption leaves us with the difficult non-convex problem

$$\inf_{\Omega_0,\ldots,\Omega_k} \left\{ \sum_{i=0}^{k} \frac{1}{2} \, |\partial \Omega_i| + \int_{\Omega_i} (c_i - I(x))^2 \, dx \right\},$$

where it is implicit that $\Omega = \cup_i \Omega_i$ and the Ω_i are pairwise disjoint. In this work, we will not consider the even more challenging problem where either the constants or number of phases are unknown. Likewise, the related methods for the multi-phase problem mentioned in the Introduction and discussed further in Section 5 also make these reasonable assumptions.

To slightly generalize our method to data terms other than the one in the Mumford-Shah functional, the multi-phase problem we will solve is

$$\inf_{\Omega_0,\ldots,\Omega_k} \left\{ \sum_{i=0}^{k} |\partial \Omega_i| + \int_{\Omega_i} g_i(x) \, dx \right\}. \tag{1}$$

For example, when the *region descriptors* $g_i(x)$ are equal to $(c_i - I(x))^2$ (where we have absorbed constant factors), we have the case of Mumford-Shah. This generalization does not add any complexity to our problem. Our assumption concerning the c_i translates to the assumption that the $g_i(x)$ are known beforehand.

When $k = 1$, expression (1) above becomes the two-phase problem

$$\inf_{\Omega_0 \subset \Omega} \left\{ |\partial\Omega_0| + \int_{\Omega_0} g_0(x)\,dx + \int_{\Omega\setminus\Omega_0} g_1(x)\,dx \right\}. \tag{2}$$

Many popular methods to solve (2) are based on the level set method (e.g., the Chan-Vese method [14]), but there is no guarantee to find a global solution. In [12], Chan et al. proposed a method for finding global minimizers through a convex formulation of the problem. They showed that (2) can be written equivalently

$$\min_{\theta\in\{0,1\}} \left\{ E(\theta) := \int_\Omega |\nabla\theta|\ dx + \int_\Omega (1-\theta(x))g_0(x) + \theta(x)g_1(x)\,dx \right\}, \tag{3}$$

where $\theta\colon \Omega \to \{0,1\}$ is a binary function that defines the segmentation: $x \in \Omega_0$ if $\theta(x) = 0$ and $x \in \Omega_1$ if $\theta(x) = 1$. However, even though $E(\theta)$ is convex, the minimization is done over a non-convex set of binary functions.

The important final step is that a relaxation may be taken to allow $\theta \in [0,1]$ without changing the minimum. Indeed, letting θ^* be any minimizer of the binary problem (3), then any $\theta \in [0,1]$ satisfies $E(\theta^*) \le E(\theta)$. Moreover, a solution of (3) can be obtained by finding a solution of the relaxed problem and then thresholding. Namely, letting now θ^* be a minimizer of the relaxed problem, the binary function $1_{\{\theta^*>t\}}$ is a minimizer of (3) for any $t \in (0,1)$. This relaxation completes the conversion of the original problem (2) to a convex minimization problem. We say the relaxation is *exact* because a true solution of the original problem can be exactly recovered from the relaxed problem.

Relaxation in this context dates back to Strang's work [31] on maximal flows. The critical property for the functional of a problem

$$\min_{x\in\{0,1\}} F(x)$$

to possess in order to have an exact relaxation to $x \in [0,1]$ is a *generalized coarea formula* of the form

$$F(x) = \int_0^1 F(1_{\{x>t\}})\,dt.$$

The article [7] contains a nice presentation of the proof that the relaxation is exact. As it turns out, the fact that the relaxation does not have a straightforward extension to vector-valued functions is the obstacle in multi-phase approaches.

3 Extension to multi-phase

In this section, we propose a method for solving the multi-phase segmentation problem (1) introduced in the previous section. We break up the discussion into three parts. First, we reformulate the problem in order to obtain a minimization problem whose unknown is an integer-valued function; next, we prove a result that allows us to solve a sequence of minimization problems whose unknown is a real-valued function; and finally, we embed the problem in a higher dimensional space to obtain a minimization problem whose objective function is convex.

3.1 Multi-phase problem representation

We return now to developing the tools to globally solve the multi-phase problem

$$\inf_{\Omega_0,\ldots,\Omega_k} \left\{ \sum_{i=0}^{k} |\partial\Omega_i| + \int_{\Omega_i} g_i(x)\, dx \right\}. \tag{4}$$

Moving forward, it becomes advantageous to represent this problem as an equivalent optimization problem over functions rather than over partitions of Ω. In [32], Vese and Chan proposed a generalization of the 2-phase Chan-Vese method that uses multiple level set functions to represent the segmentation. Later, Lie et al. [24] gave a variant of this method, which we elect to use in our approach.

We represent the partition $\Omega = \cup_i \Omega_i$ by a piecewise constant labeling function $u\colon \Omega \to \{0, \ldots, k\}$ with the property that $u = i$ on Ω_i. For convenience, we introduce the notation $\psi_i(u) := 1_{\{u=i\}}$ for the $k+1$ characteristic functions induced by u. Clearly the function u is given by $u = \sum_{i=1}^{k} i\psi_i$. The characteristic functions can further be utilized to express the boundary lengths of the segments:

$$|\partial\Omega_i| = \int_{\Omega} |\nabla\psi_i(u)|\, dx.$$

This is a consequence of the fact that the total variation of the characteristic function of a set is its perimeter. Under this representation, the segmentation problem (4) becomes the constrained optimization problem

$$\min_{u\colon \Omega\to\{0,\ldots,k\}} \left\{ \sum_{i=0}^{k} \int_{\Omega} |\nabla \psi_i(u)| + \psi_i(u) g_i(x)\, dx \right\}.$$

Next, because we eventually wish to obtain a continuous convex formulation, we relax and let $u \in \mathbb{R}$. To ensure that the function u takes values in $\{0,\ldots,k\}$, we introduce

$$K(u) = \prod_{i=0}^{k} (u-i),$$

so that if a function $u\colon \Omega \to \mathbb{R}$ satisfies identically $K(u) = 0$, then there exists a unique $i \in \{0\ldots,k\}$ for every $x \in \Omega$ such that $u(x) = i$. We use the interpolation formulas

$$\psi_i(u) = \prod_{i\neq j} \frac{(u-j)}{(i-j)},$$

which coincide with the characteristic functions on the constraint set. (In some instances, which will be clear from the context, we will use K and ψ_i to denote the corresponding functions from $\mathbb{R}$ to $\mathbb{R}$.) Our problem is now

$$\min_{u\colon \Omega\to[0,k]} \left\{ F(u) := \sum_{i=0}^{k} \int_{\Omega} |\nabla \psi_i(u)| + \psi_i(u) g_i(x)\, dx \right\} \quad \text{s.t.} \quad K(u) = 0.$$

Notice there is no harm replacing the condition $u \in \mathbb{R}$ with $u \in [0,k]$ because the constraint $K(u)$ already demands $u \in \{0,\ldots,k\}$.

3.2 An Augmented Lagrangian Method

In this subsection, we begin to describe our method for finding a global solution of

$$\min_{u\in BV(\Omega;[0,k])} F(u) \quad \text{s.t.} \quad K(u) = 0. \tag{5}$$

Note that, due to the function space over which we are minimizing, the constraint $K(u) = 0$ is in the almost everywhere (a.e.) sense. To enforce the constraint, we use an augmented Lagrangian method [4]. For any $r > 0$, let

us define the augmented Lagrangian function $L_r\colon L^2(\Omega)\times L^2(\Omega)\to[-\infty,\infty]$ by

$$L_r(u,\lambda)=F(u)+\langle\lambda,K(u)\rangle_{L^2(\Omega)}+\frac{r}{2}\left\|K(u)\right\|^2_{L^2(\Omega)}.$$

The idea is to find a sequence $\{u_j\}$ of global minimizers of

$$\min_{u\in BV(\Omega;[0,k])} L_{r_j}(u,\lambda_j) \tag{6}$$

so that for certain choices of the *multipliers* $\lambda_j\in L^2(\Omega)$ and the *penalty parameters* $r_j>0$ any limit point (in an appropriate topology) of the sequence will be a global minimizer of (5). For example, in what follows, we assume the set of multipliers $\{\lambda_j\}\subset L^2(\Omega)$ is bounded and the penalty parameters satisfy $r_1<r_2<\cdots$ and $r_j\to\infty$. In practice, other choices for λ_j and r_j can provide better convergence results; see Section 4.

We will now show that the sequence $\{u_j\}$ of global minimizers of (6) converges in L^1 (and pointwise almost everywhere) to a function $\overline{u}\in BV(\Omega)$ that is a global solution of (5). Our argument is based on the proof of [4, Prop. 2.1], which establishes the result in the finite dimensional setting.

By assumption, each function u_j is a global minimizer of $L_{r_j}(u,\lambda_j)$, which means that

$$L_{r_j}(u_j,\lambda_j)\le L_{r_j}(u,\lambda_j)\quad\text{ for all } u\in BV(\Omega;[0,k]). \tag{7}$$

Let F^* denote the optimal value of (5). Then, for any fixed $j\ge1$,

$$\begin{aligned}F^*&=\inf\left\{F(u)\colon u\in BV(\Omega;[0,k]),\ K(u)=0\right\}\\&=\inf\left\{L_{r_j}(u,\lambda_j)\colon u\in BV(\Omega;[0,k]),\ K(u)=0\right\}.\end{aligned}$$

Taking the infimum of (7) over $u\in BV(\Omega;[0,k])$ such that $K(u)=0$ yields

$$F(u_j)+\langle\lambda_j,K(u_j)\rangle_{L^2(\Omega)}+\frac{r_j}{2}\left\|K(u_j)\right\|^2_{L^2(\Omega)}\le F^*. \tag{8}$$

First we show convergence of the sequence $\{u_j\}$. To proceed, let us show that for $j=1,2,\ldots$ the sequence of integrals

$$\sum_{i=0}^{k}\int_\Omega|\nabla\psi_i(u_j)|\,dx$$

is uniformly bounded. Choosing $u \equiv 0$ in the inequality (7) yields

$$\sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(u_j)|\, dx + \frac{r_j}{2} \|K(u_j)\|^2_{L^2(\Omega)} \leq \int_\Omega |g_0(x)|\, dx + \left|\langle \lambda_j, K(u_j)\rangle_{L^2(\Omega)}\right| + \sum_{i=0}^{k} \int_\Omega |\psi_i(u_j) g_i(x)|\, dx.$$

Each term on the right-hand side is bounded by a constant independent of j, which implies that there exists $M < \infty$ such that

$$\sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(u_j)|\, dx \leq M \quad \text{for all } j. \tag{9}$$

Furthermore, since $u = \sum_{i=0}^{k} i\psi_i(u)$ for any function u, by the triangle inequality for the BV seminorm $|u|_{BV} := \int_\Omega |\nabla u|\, dx$ we have

$$|u_j|_{BV} = \left|\sum_{i=0}^{k} i\psi_i(u_j)\right|_{BV} \leq \sum_{i=0}^{k} |i\psi_i(u_j)|_{BV} \leq k \sum_{i=0}^{k} |\psi_i(u_j)|_{BV} \leq kM < \infty$$

for all $j \geq 1$. Thus, by a well-known BV compactness theorem (see, e.g., [16, p. 176]), there exists a subsequence $\{u_{n_j}\}$ converging in L^1 to $\overline{u} \in BV(\Omega)$ (which, without loss of generality, also converges a.e.).

Next, we show that $\overline{u}$ is a global minimizer of (5). Due to (9), by passing to a subsequence if necessary,

$$\lim_{j\to\infty} \sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(u_j)|\, dx \quad \text{exists},$$

and moreover, by lower semicontinuity of the BV seminorm,

$$\sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(\overline{u})|\, dx \leq \lim_{j\to\infty} \sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(u_j)|\, dx. \tag{10}$$

Since $\{\lambda_j\}$ is bounded, we may choose $\overline{\lambda} \in L^2(\Omega)$ with $\lambda_j \to \overline{\lambda}$ a.e. Continuity of K together with Lebesgue's dominated convergence theorem implies that

$$\langle \lambda_j, K(u_j)\rangle_{L^2(\Omega)} \to \langle \overline{\lambda}, K(\overline{u})\rangle_{L^2(\Omega)}.$$

Likewise, we have

$$\sum_{i=0}^{k} \int_{\Omega} \psi_i(u_j) g_i(x)\, dx \to \sum_{i=0}^{k} \int_{\Omega} \psi_i(\overline{u}) g_i(x)\, dx,$$

which, together with (10), establishes that

$$F(\overline{u}) \leq \liminf_{j\to\infty} F(u_j).$$

Consequently, taking the limit superior of (8), we have

$$F(\overline{u}) + \langle \overline{\lambda}, K(\overline{u}) \rangle_{L^2(\Omega)} + \limsup_{j\to\infty} \frac{r_j}{2} \left\| K(u_j) \right\|^2_{L^2(\Omega)} \leq F^*. \tag{11}$$

But since $\|K(u_j)\|^2_{L^2(\Omega)} \geq 0$, $r_j \to \infty$, and $F^* < \infty$, we must have $K(u_j) \to 0$ and $K(\overline{u}) = 0$, otherwise the left-hand side of (11) is infinite. It follows that $F(\overline{u}) \leq F^*$, as was to be shown.

It remains to find a global minimizer of

$$\int_{\Omega} \sum_{i=0}^{k} |\nabla \psi_i(u)| + \psi_i(u) g_i(x)\, dx + \langle \lambda_j, K(u) \rangle_{L^2(\Omega)} + \frac{r_j}{2} \left\| K(u) \right\|^2_{L^2(\Omega)} \tag{12}$$

over all $u \in BV(\Omega; [0, k])$ for fixed j. This is still quite challenging since the objective function is non-convex. The remedy is a technique for converting certain non-convex problems to convex ones, which we describe in the following subsection.

3.3 Convex relaxation by functional lifting

In [28], a method was proposed to solve minimization problems of the form

$$\min_{u\colon \Omega \to \Gamma := [\gamma_{\min}, \gamma_{\max}]} \int_{\Omega} |\nabla u| + \rho(x, u(x))\, dx,$$

where $\rho\colon \Omega \times \Gamma \to \mathbb{R}^+$ may be non-convex. The method transforms this non-convex problem into an equivalent convex problem through a change of independent variables and a relaxation on the constraints. Simply put, the idea is to reformulate the problem in terms of the super-level set function $\phi\colon \Omega \times \Gamma \to \{0, 1\}$ defined by $\phi(x, \gamma) = 1_{\{u(x) > \gamma\}}(x)$ in such a way that the

objective function of the reformulated problem is convex. The function u may then be recovered from ϕ from the layer-cake formula

$$u(x) = \gamma_{\min} + \int_{\Gamma} \phi(x, \gamma) d\gamma. \tag{13}$$

The authors of [28] call this technique *functional lifting* because we have increased the dimension of the arguments of the functions over which the minimization is taken. Indeed, rather than minimize over functions $u(x)$, we instead minimize over functions $\phi(x, \gamma)$. The idea was motivated by the discrete approach of Ishikawa [20] in which an appropriate auxiliary graph with an extended node set was created, and standard (binary) graph-cut computations (see, e.g., [22]) were used to solve the problem. We now derive a generalization of functional lifting in order to solve the minimization problem (12). In our application, we will take $\gamma_{\min} = 0$ and $\gamma_{\max} = k$ since these are the endpoints of the feasible set of (5).

One key observation is that $|\partial_\gamma \phi(x, \gamma)| = \delta(u(x) - \gamma)$. Since ϕ is the super-level set function of u,

$$\phi(x, \gamma) = \begin{cases} 1 & \text{if } \gamma < u(x), \\ 0 & \text{otherwise.} \end{cases}$$

Hence, for each $x \in \Omega$, the derivative of ϕ with respect to γ is zero except when $u(x) = \gamma$, where a jump occurs and the magnitude of this derivative is a delta function centered at $u(x) - \gamma$. Thus, for each $i = 0, \ldots, k$, we have

$$\int_{\Omega} |\nabla \psi_i(u(x))| \, dx = \int_{\Omega} \left| \nabla \left(\int_0^k \psi_i(\gamma) \partial_\gamma \phi \, d\gamma \right) \right| dx$$

For notational convenience, let

$$w_j(x, \gamma) = \sum_{i=0}^{k} (\psi_i(\gamma) g_i(x)) + \lambda_j K(\gamma) + \frac{r_j}{2} K(\gamma)^2.$$

Using the previous observation, we see that (12) is equal to

$$E_j(\phi) := \sum_{i=0}^{k} \int_{\Omega} \left| \nabla \left(\int_0^k \psi_i(\gamma) \partial_\gamma \phi \, d\gamma \right) \right| dx + \int_{\Omega} \int_0^k w_j(x, \gamma) \, |\partial_\gamma \phi| \, d\gamma \, dx. \tag{14}$$

Thus, as long as we enforce ϕ to have the super-level set property, i.e.,

$$\phi \in B := \{\phi\colon \Omega \times [0,k] \to \{0,1\}\colon \phi(x,0) = 1, \phi(x,k) = 0, \text{ and } \phi(x,\cdot) \text{ is non-increasing for each } x \in \Omega\},$$

the problem (12) is equivalently

$$\min_{\phi \in B} E_j(\phi). \tag{15}$$

We then use (13) to obtain u from ϕ. Observe that the objective function E_j is convex in ϕ because the non-convex functions ψ_i and K do not depend on ϕ; the non-convexity has somehow been integrated out.

Finally, we perform a relaxation on ϕ to allow $\phi \in [0,1]$. We let

$$R := \{\phi\colon \Omega \times [0,k] \to [0,1]\colon \phi(x,0) = 1, \phi(x,k) = 0, \text{ and } \phi(x,\cdot) \text{ is non-increasing for each } x \in \Omega\},$$

and consider the convex minimization problem

$$\min_{\phi \in R} E_j(\phi). \tag{16}$$

The idea is to compute a minimizer ϕ^* of (16) and threshold to obtain a function $1_{\{\phi > t\}} \in B$, following the ideas of [12] mentioned in Section 2. In order to guarantee that $1_{\{\phi > t\}}$ is a solution to (15), we need the generalized coarea formula

$$E_j(\phi) = \int_0^1 E_j(1_{\{\phi > t\}})\, dt. \tag{17}$$

As it happens, there exist $\phi \in R$ such that (17) fails. However, in practice, we find that minimizers of (16) satisfy (17). We will examine this formula further in Section 5, where we give experimental evidence that the formula holds for minimizers, but also give a counterexample to show that does not hold in general.

We summarize the results of the section with the following theorem.

Theorem 1. *Suppose* $(\{\lambda_j\}, \{r_j\})_{j=1}^{\infty}$ *is such that* $\{\lambda_j\}$ *is bounded,* $r_1 < r_2 < \cdots$, *and* $r_j \to \infty$. *For each* j, *globally solve the convex minimization problem*

$$\min_{\phi \in R} E_j(\phi)$$

to obtain a function ϕ_j^*. Suppose that ϕ_j^* satisfies the generalized coarea formula

$$E_j(\phi_j^*) = \int_0^1 E_j(1_{\{\phi_j^* > t\}})\, dt.$$

Then for any $t \in (0,1)$, let

$$u_j(x) = \int_0^k 1_{\{\phi_j^* > t\}}(x,\gamma) d\gamma.$$

Then the sequence of functions u_j converges to a global solution u^* of

$$\min_{u \in BV(\Omega;[0,k])} \left\{ \sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(u)| + \psi_i(u) g_i(x)\, dx \right\} \quad s.t. \quad K(u) = 0.$$

Equivalently, this provides a global solution to the multi-phase segmentation problem

$$\inf_{\Omega_0,\ldots,\Omega_k} \left\{ \sum_{i=0}^{k} |\partial \Omega_i| + \int_{\Omega_i} g_i(x)\, dx \right\}$$

if we put $\Omega_i = \{x \in \Omega \colon u^*(x) = i\}$.

Even when each ϕ_j^* does not satisfy the generalized coarea formula, we can still apply our method to solve the multi-phase segmentation problem, although we cannot guarantee we have found a global solution. In the next section, we implement the algorithm and the examine the extent to which we can find global solutions for particular test images.

4 Algorithm and Numerical Results

We now describe an algorithm to solve the convex minimization problem

$$\min_{\phi \in R} E_j(\phi),$$

and use what we have developed in the previous section to provide solutions to (1). Our objective is simply to describe a global algorithm resulting from our theory rather than developing the most efficient algorithm possible, which we leave for future work. Recall that

$$E_j(\phi) = \sum_{i=0}^{k} \int_\Omega \left| \nabla \left(\int_0^k \psi_i(\gamma) \partial_\gamma \phi\, d\gamma \right) \right| dx + \int_\Omega \int_0^k w_j(x,\gamma)\, |\partial_\gamma \phi|\, d\gamma\, dx$$

with

$$w_j(x,\gamma) = \sum_{i=0}^{k} (\psi_i(\gamma) g_i(x)) + \lambda_j K(\gamma) + \frac{r_j}{2} K(\gamma)^2.$$

Using the dual formulation of TV, this can be written

$$\max_{\mathbf{p}\in X} \int_\Omega \int_0^k \sum_{i=0}^{k} -\operatorname{div} p_i(x,\gamma)\psi_i(\gamma)\, |\partial_\gamma \phi| + w_j(x,\gamma)\, |\partial_\gamma \phi|\; d\gamma\, dx,$$

where

$$X = \{\mathbf{p} = (p_0, \ldots, p_k) \colon \Omega \times [0,k] \to \mathbb{R}^{d\times(k+1)} \colon\ \|p_i\|_{L^\infty} \leq 1 \quad \forall i = 0, \ldots, k\}.$$

Thus, we now have the min-max problem

$$\min_{\phi\in R} \max_{\mathbf{p}\in X} \left\{ \underbrace{\int_\Omega \int_0^k \sum_{i=0}^{k} -\operatorname{div} p_i(x,\gamma)\psi_i(\gamma)\, |\partial_\gamma \phi| + w_j(x,\gamma)\, |\partial_\gamma \phi|\; d\gamma\, dx}_{\Phi_j(\phi,\mathbf{p})} \right\}.$$

Note that we have an optimization problem over the set $R \times X$, where R and X are compact, convex subsets of linear topological spaces, such that $\Phi_j(\phi, \cdot)$ is concave on X for all $\phi \in R$ and $\Phi_j(\cdot, \mathbf{p})$ is convex on R for all $\mathbf{p} \in X$. By Sion's minimax theorem [30], the minimization and maximization operations may be interchanged.

For the inner maximization step, observe that the terms dependent on the dual variables p_i decouple. In other words, to conduct the maximization, we may separately solve for each i:

$$\max_{|p_i(x,\gamma)|\leq 1} \left\{ \int_\Omega \int_0^k -\operatorname{div} p_i(x,\gamma)\psi_i(\gamma)\, |\partial_\gamma \phi|\; d\gamma\, dx \right\},$$

This may be solved using the iterative scheme [10]:

$$p_i^{n+1}(x,\gamma) = \mathcal{P}_X \left(p_i^n(x,\gamma) + \tau_p \psi_i(\gamma) \nabla \left(|\partial_\gamma \phi^n| \right) \right), \tag{18}$$

where τ_p denotes the step size of updates for the p_i variables. The operator $\mathcal{P}_X$ is the projection onto the set X, i.e.,

$$(\mathcal{P}_X(\mathbf{q}))(x,\gamma) = \left(\frac{\mathbf{q}_0(x,\gamma)}{\max(|\mathbf{q}_0(x,\gamma)|, 1)}, \ldots, \frac{\mathbf{q}_k(x,\gamma)}{\max(|\mathbf{q}_k(x,\gamma)|, 1)} \right)$$

for all $\mathbf{q}\colon \Omega \times [0,k] \to \mathbb{R}^{d\times(k+1)}$.

For the outer minimization problem, we use an explicit gradient descent method derived from the corresponding Euler-Lagrange equation. Again, let us emphasize that this is just one method to perform this minimization subproblem and certainly not the most efficient. It was instead chosen for ease of implementation. Put

$$W_j(x,\gamma,\mathbf{p}) = \sum_{i=0}^{k} -\operatorname{div} p_i(x,\gamma)\psi_i(\gamma) + w_j(x,\gamma).$$

The problem is then, for fixed $\mathbf{p}$,

$$\min_{\phi\in R}\left\{\int_\Omega\int_0^k W(x,\gamma,\mathbf{p})\,|\partial_\gamma\phi|\;d\gamma\,dx\right\}.$$

Introducing an artificial time t, the gradient descent PDE is

$$\frac{\partial\phi}{\partial t} = -\nabla_\gamma\cdot\left(W(x,\gamma,\mathbf{p})\frac{\partial_\gamma\phi}{|\partial_\gamma\phi|}\right).$$

We write the divergence operator $\nabla_\gamma\cdot = \operatorname{div}_\gamma$ instead of ∂_γ since these will be different in the finite difference numerical scheme. To avoid the degenerate case when $\partial_\gamma\phi = 0$, we regularize this equation by some small $\varepsilon > 0$ to yield

$$\frac{\partial\phi}{\partial t} = -\nabla_\gamma\cdot\left(W(x,\gamma,\mathbf{p})\frac{\partial_\gamma\phi}{\sqrt{\varepsilon^2+(\partial_\gamma\phi)^2}}\right).$$

This gives the iterative scheme

$$\phi^{n+1} = \mathcal{P}_R\left\{\phi^n + \tau_\phi\nabla_\gamma\cdot\left(W(x,\gamma,\mathbf{p})\frac{\partial_\gamma\phi^n}{\sqrt{\varepsilon^2+(\partial_\gamma\phi^n)^2}}\right)\right\}, \tag{19}$$

where $\mathcal{P}_R$ is the Euclidean projection onto the convex set R. In practice, it turns out that the condition that $\phi(x,\dot{})$ is non-increasing for each x is automatically satisfied, although we have no proof of this property. This is related to an ordering property established in [11, Prop. 4.3]. In this case, the projection on R reduces to a projection onto the set where $\phi(x,0) = 1$ and $\phi(x,k) = 0$ for all x, which is a simple truncation operation.

To simplify the discussion of the numerical implementation of the algorithm, we consider only the $d = 2$ case and let $\Omega = (0,1)^2$. We use the standard discretization in the spatial domain

$$\Omega^h = \{1, \ldots, N\} \times \{1, \ldots, N\}$$

with spatial step size $h = 1/N$, and use the discretization

$$\Gamma^{\Delta\gamma} = \{(g-1)\Delta\gamma \colon 1 \leq g \leq N_\gamma\}$$

with $N_\gamma = 1 + k/\Delta\gamma$. A spatially continuous function $f \colon \Omega \times \Gamma \to \mathbb{R}$ is now approximated by a discrete function $f^{h,\Delta\gamma}$, from which we will often omit the superscripts. For $f : \{1, \ldots, N\}^2 \times \{1, \ldots, N_\gamma\} \to \mathbb{R}$, we use the notation $f_{i,j,g} := f(i,j,g)$. We have the following finite difference formulas for the derivative operators:

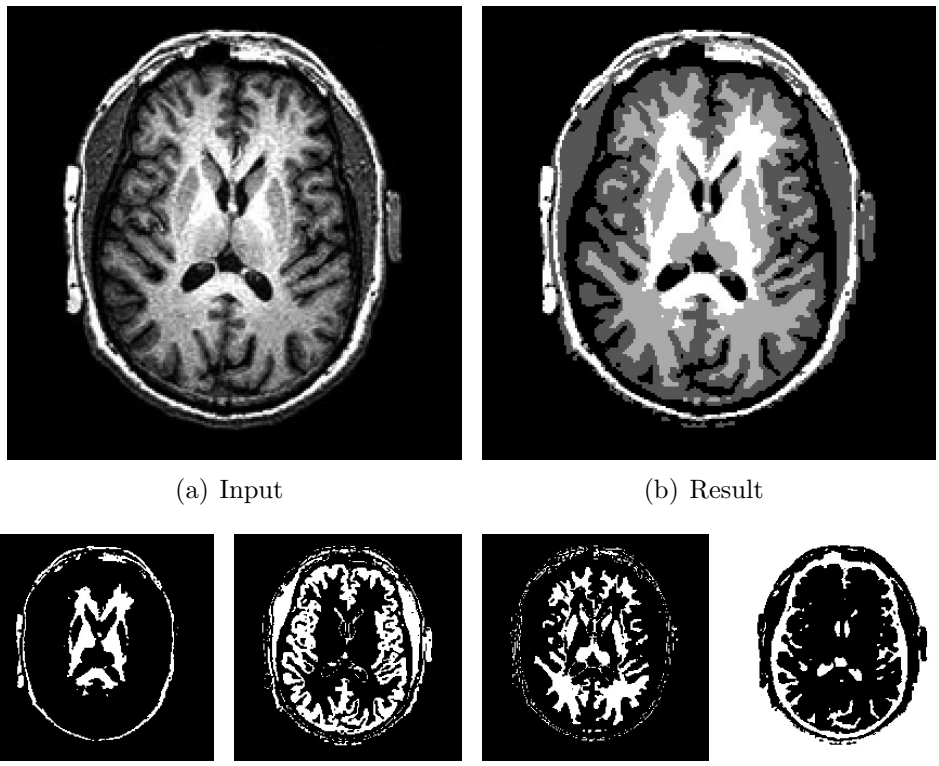
$$\begin{cases} (\operatorname{div}_\gamma f)_{i,j,g} & = (f_{i,j,g} - f_{i,j,g-1})/\Delta\gamma \\ (\partial_\gamma f)_{i,j,g} & = (f_{i,j,g+1} - f_{i,j,g})/\Delta\gamma \\ (\nabla f)^1_{i,j,g} & = (f_{i+1,j,g} - f_{i,j,g})/h \\ (\nabla f)^2_{i,j,g} & = (f_{i,j+1,g} - f_{i,j,g})/h \end{cases}$$

with appropriate boundary conditions (see [9]). In our experiments, we are given a discrete image defined on Ω^h and let $h = 1$. We set $\Delta\gamma = 0.25$.

We mentioned in the previous section that the augmented Lagrangian variables λ^j and r^j should satisfy the restrictions of Theorem 1 in order to guarantee that the algorithm obtains a globally optimal solution. In practice, these variables may be chosen differently to increase the efficiency of the algorithm. For example, in the augmented Lagrangian technique of [24], the penalty parameters r^j are taken to be equal to some constant r for all j, and the Lagrange multipliers are updated as $\lambda^{j+1} = \lambda^j + rK(u^j)$; this is often referred to as the method of multipliers [4]. We elect to use this update in our implementation. We let $r = 1$ and initialize $\lambda^0 = 1$.

We first illustrate our algorithm on two real-world examples. Figure 1(a) shows an MRI brain image common to medical imaging applications. Figure 1(b) shows the result with four phases. We show each phase in a Figures 1(c)–(f).

We also apply the algorithm to a color image in Figure 2 using four phases. For color images, where the image is a vector $\overrightarrow{I}$, the region descriptors $g_i(x)$ can be defined as $\left\| \overrightarrow{I}(x) - \overrightarrow{c}_i \right\|^2$, where $\overrightarrow{c}_i$ is the vector mean in phase i.



(a) Input (b) Result

(c) Phase 1 (d) Phase 2 (e) Phase 3 (f) Phase 4

Figure 1: Segmentation of an MRI brain image into four phases.

(a) Input (b) Result

Figure 2: Segmentation of a natural color image into four phases.

Lastly, we apply our algorithm to the well known synthetic triple junction example in Figure 3. We see that our segmentation algorithm successfully finds the optimal solution.

We should point out that the method of Pock et al. [27, 11] is also able to find the correct solution, where all three interfaces meet with an angle of 120°. The methods of [33, 23] do not initially provide the correct segmentation for the triple junction, but can compute the desired result when a procedure similar to thresholding is conducted; see Section 5.2.

Recall that our method assumes that the region descriptors (which, in the case of piecewise constant Mumford-Shah, are characterized by the mean intensities of each segment) are known *a priori*. In our experiments, we apply a k-means clustering algorithm to determine these values. Variations of our algorithm could include updating these values during the segmentation process. An alternative would be to run the entire algorithm several times, updating these values after convergence, to yield slightly better results.

5 Discussion

In the previous sections, we described a method for globally solving the multi-phase segmentation problem under the Mumford-Shah/Potts model. We now discuss a convex relaxation method from the literature and exhibit

(a) Input (b) Result

Figure 3: Segmentation of a synthetic image. The optimal solution is given by a 120° triple junction.

a relationship with our approach.

5.1 Convex relaxation over the simplex

Recall that we decompose the labeling function $u\colon \Omega \to \{0, \dots, k\}$ with

$$\psi_i = \begin{cases} 1 & \text{if } u = i, \\ 0 & \text{otherwise.} \end{cases}$$

This means that $u = \sum_i i\psi_i$. Rather than proceed as in Section 3, consider instead the formulation

$$\min_{\mathbf{\Psi}=(\psi_0,\dots,\psi_k)\in S} \left\{ F(\Psi) := \sum_{i=0}^{k} \int_\Omega |\nabla \psi_i| + \psi_i g_i(x)\, dx \right\}, \tag{20}$$

where

$$S = \left\{ (f_0, \dots, f_k)\colon \Omega \to \{0,1\}^{k+1}\colon \sum_{i=0}^{k} f_i(x) = 1 \text{ for all } x \in \Omega \right\}.$$

The natural multi-dimensional analogue of the one-dimensional convex relaxation approach (like that of [12] or [28]) would be to minimize over the

convex probability simplex

$$C = \left\{ (f_0, \dots, f_k) \colon \Omega \to [0,1]^{k+1} \colon \sum_{i=0}^{k} f_i(x) = 1 \text{ for all } x \in \Omega \right\}$$

so that we have the convex minimization problem

$$\min_{(\psi_0,\dots,\psi_k)\in C} \left\{ \sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(x)| + \psi_i(x) g_i(x)\, dx \right\}. \tag{21}$$

This is essentially the method proposed in [33]. To solve (21), additional variables $(v_0, \dots, v_k) \colon \Omega \to \mathbb{R}^k$ were introduced to decouple the regularization and data terms [1, 6], yielding

$$\min_{(\psi_0,\dots,\psi_k),(v_0,\dots,v_k)\in C} \left\{ \int_\Omega \sum_i |\nabla \psi_i| + \sum_i \frac{1}{2\mu}(\psi_i - v_i)^2 + \sum_i \psi_i g_i\, dx \right\},$$

where $\mu > 0$ is a parameter that controls the quadratic approximation of ψ_i and v_i for each i. Similarly, Lellmann et al. [23] considered the relaxed problem

$$\min_{(\psi_0,\dots,\psi_k)\in C} \left\{ \int_\Omega \sqrt{\sum_i |\nabla \psi_i|^2} + \psi_i g_i(x)\, dx \right\}. \tag{22}$$

Up to constant factors, this is equivalent to (21) when $(\psi_0, \dots, \psi_k) \in S$. To solve the convex optimization problem (22), a Douglas-Rachford splitting algorithm [15] was used.

The issue with both of these methods is that the minimizers of (21) and (22) cannot be guaranteed to lie in the set S. More precisely, if $\Psi^* = (\psi_0^*, \dots, \psi_k^*)$ is a minimizer of, for example, (21), then it could be the case $F(\Psi^*)$ is strictly less than the minimum of F over S. It is unfortunately not possible (in contrast to [12] or [28]) to claim that a minimizer of (20) is given by constructing $(1_{\{\psi_0^* > t\}}, \dots, 1_{\{\psi_k^* > t\}})$ for any $t \in (0,1)$, since this thresholded vector need not even be in S. Instead, one could consider assigning, for each $x \in \Omega$, the unit vector $\vec{\mathbf{e}}_{i*}$ with $i^* = \operatorname{argmax}_j \psi_j(x)$. However, the authors of [33] and [23] have no proof that such an assignment would remain a minimizer. It is not clear whether there is a concept like the generalized coarea formula in this setting which could ensure the obtained binary vector would globally solve (21). In short, it is not clear whether this relaxation

is exact and therefore it cannot guarantee a global solution of the original problem.

We should mention the related work of Bae et al. [3] which considered the primal-dual formulation of (21). Using the dual formulation of total variation [13, 8, 9], the minimization problem (21) can be written as the saddle point problem

$$\min_{\Psi \in C} \max_{|p_i| \leq 1} \left\{ \sum_{i=0}^{k} \int_{\Omega} \psi_i(x) \left(\operatorname{div} p_i(x) + g_i(x) \right) \, dx \right\}. \tag{23}$$

By Sion's minimax theorem [30], the minimum and maximum operations can be interchanged to obtain

$$\max_{|p_i| \leq 1} \min_{\Psi \in C} \left\{ \sum_{i=0}^{k} \int_{\Omega} \psi_i(x) \left(\operatorname{div} p_i(x) + g_i(x) \right) \, dx \right\}.$$

It is easy to verify that this reduces to

$$\max_{|p_i| \leq 1} \int_{\Omega} \min \left(\operatorname{div} p_0 + g_0, \ldots, \operatorname{div} p_k + g_k \right) \, dx.$$

Let $\mathbf{p}^*$ be a maximizer of (5.1) and Ψ^* be a maximizer of (21) such that $(\Psi^*, \mathbf{p}^*)$ is an optimal pair of (23). If, for each $x \in \Omega$, the values $\operatorname{div} p_0 + g_0, \ldots, \operatorname{div} p_k + g_k$ have a unique minimum, then $\Psi^* \in C$ is necessary binary. Thus, under these assumptions, solving the convex relaxation (21) solves the original multi-phase segmentation problem (1).

5.2 Connection with our method

Our method solves a sequence of minimization problems of the form

$$\min_{\phi \colon \Omega \times [0,k] \to [0,1]} \left\{ E_j(\phi) \colon \phi(x,0) = 1, \phi(x,k) = 0, \phi(x,\cdot) \text{ non-increasing } \forall x \right\}.$$

Observe that at each $x \in \Omega$, the constraints on the function ϕ imply that the quantity $|\partial_\gamma \phi(x,\cdot)|$ is a probability measure on the interval $[0,k]$. Using ideas like those discussed in Section 3.2, one can show that the sequence $\{|\partial_\gamma \phi_j|\}$ converges in $L^1(\Omega \times [0,k])$, and that the limiting family of probability measures

$$\left\{ \nu_x := \left| \partial_\gamma \overline{\phi}(x,\cdot) \right| \right\}_{x \in \Omega}$$

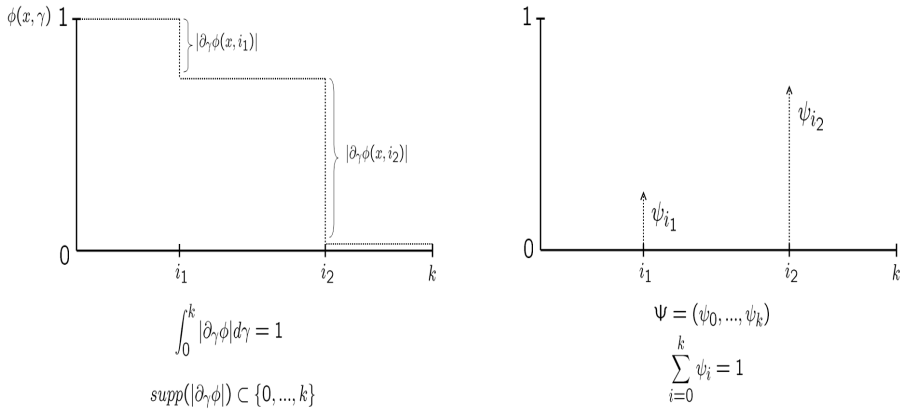


Figure 4: An example of a graphical representation showing the correspondence between a function $\phi(x,\cdot)$ and the $(k+1)$-tuple $(\psi_0(x),\ldots,\psi_k(x))$ for fixed $x \in \Omega$.

has support only on the integer set $\{0,\ldots,k\}$. When the generalized coarea formula holds, we may threshold to obtain a measure that, in addition, has support at exactly one integer for each x; this integer is the label $u(x)$.

Let us consider instead the measure ν_x before thresholding. Of course, this ignores the crucial fact that thresholding is necessary to obtain an equivalent minimization problem via functional lifting. Because ν_x is a probability measure with integer support, there is a clear one-to-one correspondence between the $(k+1)$-tuples $(\nu_x(0),\ldots,\nu_x(k))$ in our setting and $(u_0(x),\ldots,u_k(x))$ in that of Section 5.1. In each case, the ith component conveys the probability of the point x having the label i. See Figure 4.

The family of measures ν_x is a special case of more general concept known as *parametrized* (or *Young*) *measures* [26]. Consider a sequence of functions $f_j : \Omega \to \mathbb{R}$ that converge to f weakly* in L^∞, i.e.,

$$\int_\Omega g f_j \, dx \to \int_\Omega g f \, dx$$

for all $g \in L^1(\Omega)$, and let $\varphi\colon \mathbb{R} \to \mathbb{R}$ be continuous. Then the sequence $\varphi(f_j)$ also converges weakly* in L^∞, but in general not to $\varphi(f)$. Instead, the limit is

$$\overline{\varphi}(x) := \int_{\mathbb{R}} \varphi(\gamma) \, d\nu_x(\gamma).$$

Thus, we can think of the parametrized measure as the limiting probability distribution of the values of the f_j. The situation at hand is far less troublesome because we have strong convergence to a function satisfying our constraint $K(u) = 0$, which guarantees integer support.

Thus, while the methods in [33, 23] use an explicit convex simplex constraint, i.e., $\sum_i u_i = 1$, our method has an implicit simplex constraint of the magnitude of the limiting probability measure ν_x. The sequence of minimizers ϕ_j^* of the problems $\min_{\phi \in R} E_j(\phi)$ converge to a function ϕ^* with the property that $|\partial_\gamma \phi^*(x,i)| = \nu_x(i)$ and $\sum_i \nu_x(i) = 1$. The advantage is that, using our representation, we can use thresholding to obtain solutions that provide a well-defined labeling for each $x \in \Omega$. The generalized coarea formula can be examined to determine whether this labeling is optimal. In contrast, thresholding a vector $(u_0, \ldots, u_k)$ such that $\sum_i u_i = 1$ does not necessarily yield a vector in the simplex. In the setting of [33, 23], there is no apparent analogue of the coarea formula that allows us to show the solution obtained is optimal.

5.3 The generalized coarea formula

In the final part of this section, we examine in more detail the important formula

$$E_j(\phi) = \int_0^1 E_j(1_{\{\phi > t\}})\, dt, \tag{24}$$

which we have referred to throughout as the generalized coarea formula. Recall the expression for E_j defined in (14), which we repeat here for convenience:

$$E_j(\phi) = \sum_{i=0}^{k} \int_\Omega \left| \nabla \left(\int_0^k \psi_i(\gamma) \partial_\gamma \phi \, d\gamma \right) \right| \, dx + \int_\Omega \int_0^k w_j(x,\gamma) \, |\partial_\gamma \phi| \, d\gamma \, dx.$$

By the standard coarea formula, it follows quickly that

$$\int_\Omega \int_0^k w_j(x,\gamma) \, |\partial_\gamma \phi| \, d\gamma \, dx = \int_0^1 \int_\Omega \int_0^k w_j(x,\gamma) \left| \partial_\gamma 1_{\{\phi > t\}} \right| \, d\gamma \, dx \, dt.$$

It suffices to consider the reduced formula

$$\int_\Omega \left| \nabla \left(\int_0^k \psi(\gamma) \partial_\gamma \phi \, d\gamma \right) \right| \, dx = \int_0^1 \int_\Omega \left| \nabla \left(\int_0^k \psi(\gamma) \partial_\gamma 1_{\{\phi > t\}} \, d\gamma \right) \right| \, dx \, dt. \tag{25}$$

Consider the situation where, for each $x \in \Omega$, the function ϕ is a decreasing staircase function like the top part of Figure 4. In other words, we have that $\partial_\gamma \phi(x,\gamma) = -\sum_{j=0}^{k} u_j(x)\delta(\gamma - j)$ with $\sum_{j=0}^{k} u_j(x) = 1$ for all $x \in \Omega$. Furthermore, let $\psi(x) = \psi_i(x)$ for some $i \in \{0, 1, \ldots, k\}$, so that $\psi_i(j) = \delta_{ij}$ for integer j. It follows that

$$\int_0^k \psi_i(\gamma)\partial_\gamma \phi \, d\gamma = \int_0^k \psi_i(\gamma) \left(-\sum_{j=0}^{k} u_j(x)\delta(\gamma - j) \right) d\gamma = -u_i(x).$$

Thus, the left-hand side of (25) becomes

$$\int_\Omega \left| \nabla \left(\int_0^k \psi_i(\gamma)\partial_\gamma \phi \, d\gamma \right) \right| dx = \int_\Omega |\nabla u_i(x)| \, dx = \int_0^1 \int_\Omega \left| \nabla 1_{\{u_i > t\}} \right| dx \, dt, \tag{26}$$

where the last equality follows from the usual coarea formula.

On the other hand, our assumption on ϕ implies that $1_{\{\phi > t\}}$ is a step function whose jump occurs at the integer j such that $t \in A_j$ with

$$A_j = \left[\sum_{m>j} u_m, \sum_{m \geq j} u_m \right).$$

In other words, we partition the interval $[0, 1)$ into disjoint intervals $A_0, \ldots, A_k$ determined by $u_0, \ldots, u_k$, and jump of the step function $1_{\{\phi > t\}}$ occurs at the integer j corresponding to the index of the interval in which t lies. This implies that

$$\int_0^k \partial_\gamma 1_{\{\phi > t\}} \psi_i(\gamma) \, d\gamma = \int_0^k -\delta(j - \gamma)\psi_i(\gamma) = \begin{cases} -1 & t \in A_i, \\ 0 & \text{else.} \end{cases}$$

Consequently, the right-hand side of (25) is simply

$$\int_0^1 \int_\Omega |\nabla 1_{A_i}| \, dx \, dt. \tag{27}$$

Consider the following example, which will show that (26) and (27) are in general not equal, thereby proving that the generalized coarea formula (24) cannot hold for all ϕ. Suppose that $\Omega = [0, 3)$ and put $\Omega_i = [i, i + 1)$ for $i = 0, 1, 2$. Let the $u = (u_0, u_1, u_2)$ be piecewise constant on each Ω_i with values given by the values in Table 1.

	u_0	u_1	u_2
Ω_0	1/2	1/3	1/6
Ω_1	1/6	1/2	1/3
Ω_2	1/3	1/6	1/2

Table 1: Counterexample for generalized coarea formula

Let us examine (26) and (27) for $i = 1$. In this case, the plot of $1_{\{u_1 > t\}}$ is shown in Figure 5(a), and we see that the quantity (26) is given by the sum of the magnitudes of the jumps of u_1, which equals $1/6+1/3 = 1/2$. In contrast, the plot of 1_{A_1} is shown in Figure 5(b). To compute (27), we sum the magnitudes of the symmetric differences across the boundaries, $|A_1(\Omega_0)\Delta A_1(\Omega_1)|$ and $|A_1(\Omega_1)\Delta A_1(\Omega_2)|$. Recall the symmetric difference between two sets A and B is given by

$$A\Delta B = (A \cup B) \setminus (A \cap B).$$

Observe that

$$A_1(\Omega_0) = [1/6, 1/2),\ A_1(\Omega_1) = [1/3, 5/6),\ A_1(\Omega_2) = [1/2, 2/3).$$

The sum of the symmetric differences across boundaries is thus $1/2 + 1/3 = 5/6$. We have therefore constructed a counterexample to the generalized coarea formula.

Nevertheless, there is reason to believe that for minimizers of E_j satisfy the generalized coarea formula. In practice, these minimizers look like functions of the form

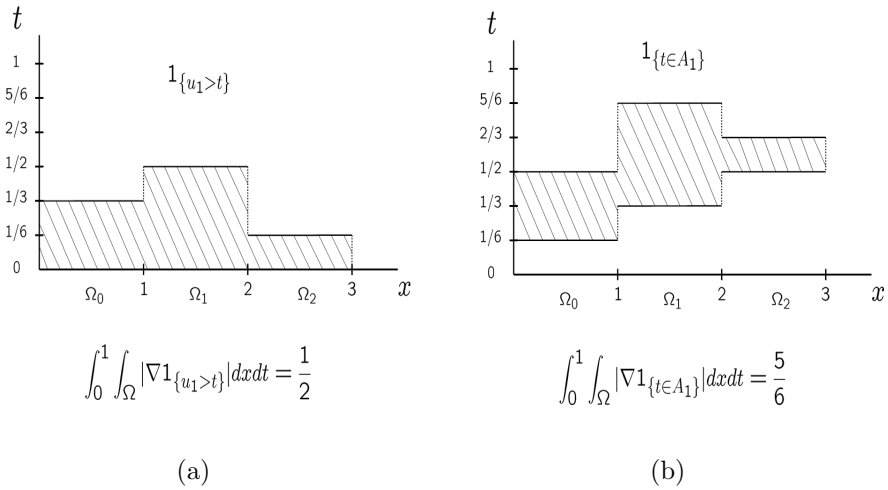
$$\phi(x, \gamma) = 1 - H_\epsilon(\gamma - \alpha(x)), \tag{28}$$

where

$$H_\epsilon(z) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{z}{\epsilon}\right)\right)$$

is the well known regularized approximation to the Heaviside function which converges to $H(z)$ as $\epsilon \to 0^+$. The function $\alpha(x)$ is simply any function of x, e.g., a labeling function or the image itself, that takes values in the interval $[0, k]$.

In Table 2, we show the numerical values of the two sides of the reduced expression (25) for various choices of the regularization parameter ϵ and the



(a) (b)

Figure 5: A counterexample to the generalized coarea formula.

function ψ. In these experiments, we put $\alpha(x)$ equal to the image $I(x)$ for the brain image considered in Section 4. We see that these expressions are approximately equal to each other, giving reason to believe that the generalized coarea formula is often satisfied in practice.

When the generalized coarea formula is approximately satisfied, we can make precise statements about the approximation of the obtained solution to the global minimizer of the original problem.

Proposition 2. *Let $\epsilon > 0$. Suppose that for $j = 1, 2, \ldots$, we have the bound*

$$\left| E_j(\phi_j^*) - \int_0^1 E_j(1_{\{u_j^*>t\}})\, dt \right| < \epsilon, \tag{29}$$

where ϕ^* is the global minimizer of E_j over $\phi \in R$. Then for any $t \in [0, 1)$ the functions

$$u_j := \int_0^k 1_{\{\phi_j^*>t\}}\, d\gamma$$

converge to a function $\tilde{u}$ such that $K(\tilde{u}) = 0$ and

$$|F(\tilde{u}) - F^*| < \epsilon,$$

where F^* is the optimal value of F.

Table 2: Numerical values of the left and right-hand sides of (25) for several choices of ψ with ϕ of the form (28).

$\psi(\gamma)$	ϵ	LHS	RHS
$\cos(\gamma)$	0.1	10.84	10.84
	1	5.08	5.12
	2	8.02	8.07
$\frac{(\gamma-2)(\gamma-3)}{2}$	0.1	23.93	23.93
	1	11.48	11.60
	2	17.95	18.07
γ^3	0.1	3.44	3.44
	1	1.58	1.61
	2	2.44	2.48
$\sqrt{\lvert\gamma\rvert}$	0.1	0.09	0.09
	1	0.11	0.12
	2	0.14	0.15

Proof. The proof uses many of the ideas from Section 3.2 and we follow same the notation. The bound (29) implies that $E_j(1_{\{\phi_j^*>t\}}) < E_j(\phi^*) + \epsilon$. Since $1_{\{\phi_j^*>t\}} \in B$, we may define

$$u_j' = \int_0^k 1_{\{\phi_j^*>t\}}\, d\gamma$$

and have $E_j(1_{\{\phi_j^*>t\}}) = L_{r_j}(u_j', \lambda_j)$. Moreover, $L_{r_j}(u_j', \lambda_j) < L_{r_j}(u_j, \lambda_j) + \epsilon$, where u_j is a global minimizer of $L_{r_j}(\cdot, \lambda_j)$. A similar argument to Section 3.2 shows that any limit point of the functions u_j', say $\tilde{u}$, satisfies the constraint $K(\tilde{u}) = 0$. Since we know that $K(u^*) = 0$ and $F^* = F(u^*)$, the desired bound follows. □

6 Conclusion

In this paper, we have proposed a method for solving the piecewise constant Mumford-Shah image segmentation problem

$$\inf_{\Omega_0,\ldots,\Omega_k} \left\{ \sum_{i=0}^{k} \frac{1}{2} |\partial\Omega_i| + \int_{\Omega_i} (c_i - I(x))^2 \, dx \right\},$$

in which $k \in \mathbb{Z}^+$ and $c_0, \ldots, c_k \in \mathbb{R}$ are fixed. First, we proved that we can globally solve the problem via a sequence of non-convex minimization problems of a specific form. Our method embeds each subproblem into a higher dimensional space to obtain a convex objective functional, and then relaxes the constraint set to obtain a convex minimization problem that can be solved globally. We cannot prove each subproblem can be solved globally because a generalized coarea formula does not hold, which means the relaxation in not exact. However, we provide empirical evidence that shows the formula holds in practice, and demonstrate the effectiveness of our method on several image segmentation examples.

In the discrete setting, this problem is known as the Potts model and has been shown to be NP-hard. As a result, we cannot expect to compute global solutions to the problem in all cases. Instead, our method adds to the existing literature of convex relaxation methods that provide approximate global solutions. In addition, we provide a connection between our method and the convex relaxation method of [33, 23] that allows us to estimate the extent to which global solutions are found via the coarea formula.

Acknowledgments

The authors would like to thank A. Chambolle for helpful discussion. In particular, this helped the authors develop the material in Sections 5.2-5.3.

References

[1] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition–modeling, algorithms, and parameter selection. *Int. J. Comput. Vision*, 67(1):111–136, 2006.

[2] E. Bae and X.-C. Tai. Efficient global optimization for the multiphase chan-vese model of image segmentation by graph cuts. In *EMMCVPR*, 2009.

[3] E. Bae, J. Yuan, and X.-C. Tai. Global minimization for continuous multiphase partitioning problems using a dual approach. In *UCLA CAM Report 09-75*, September, 2009.

[4] D. P. Bertsekas. *Constrained optimization and Lagrange Multiplier methods*. 1982.

[5] Y. Boykov, O. Veksler, and R. Zabih. A new algorithm for energy minimization with discontinuities. In *EMMCVPR '99: Proceedings of the Second International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 205–220, London, UK, 1999. Springer-Verlag.

[6] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *J. Math. Imaging Vis.*, 28(2):151–167, 2007.

[7] M. Burger and M. Hintermüller. Projected gradient flows for BV / level set relaxation. *PAMM*, 5:11–14, 2005.

[8] J. Carter. *Dual methods for total variation–based image restoration*. PhD thesis, University of California, Los Angeles (Advisor: T.F. Chan), 2001.

[9] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004. Special issue on mathematics and image analysis.

[10] A. Chambolle. Total variation minimization and a class of binary mrf models. In *EMMCVPR*, pages 136–152, 2005.

[11] A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Technical report TR-2008-05, Dept. of Computer Science, University of Bonn, Bonn, Germany, November 2008.

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

[13] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.

[14] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb 2001.

[15] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439, 1956.

[16] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. 2000.

[17] L. Ford and D. Fulkerson. *Flows in Networks*. 1962.

[18] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[19] 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*, 51(2):271–279, 1989.

[20] H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, 2003.

[21] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. 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.

[22] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):147–159, 2004.

[23] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Technical report, IWR, University of Heidelberg, October 2008.

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

[25] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[26] P. Pedregal. *Parametrized measures and variational principles*. 1997.

[27] T. Pock, A. Chambolle, H. Bischof, and D. Cremers. A convex relaxation approach for computing minimal partitions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, 2009.

[28] 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.

[29] R. B. Potts and C. Domb. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48:106–109, 1952.

[30] M. Sion. On general minimax theorems. *Pacific J. Math.*, 8:171–176, 1958.

[31] G. Strang. Maximal flow through a domain. *Math. Programming*, 26(2):123–143, 1983.

[32] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International Journal of Computer Vision*, 50:271–293, 2001.

[33] 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.