Convex Formulation and Exact Global Solutions for Multi-phase Piecewise Constant Mumford-Shah Image Segmentation

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

August 20, 2009

Abstract

Most variational models for multi-phase image segmentation are non-convex and possess multiple local minima, which makes solving for a global solution an extremely difficult task. In this work, we provide a method for computing a global solution for the (non-convex) multi-phase piecewise constant Mumford-Shah (spatially continuous Potts) image segmentation problem. Our approach is based on using a specific representation of the problem due to Lie et al. [27]. We then rewrite this representation using the dual formulation for total variation so that a variational convexification technique due to Pock et al. [30] may be employed. Unlike some recent methods in this direction, our method can guarantee that a global solution is obtained. We believe our method to be the first in the literature that can make this claim. 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 examples, including a medical imaging application.

*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 ONR N00014-03-1-0071 and ONR MURI subcontract from Stanford University.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2009		**00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Convex Formulation and Exact Global Solutions for Multi-phase Piecewise Constant Mumford-Shah Image Segmentation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Mathematics,University of California,Los Angeles,CA**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**27**	

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 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 [21]. Greig et al. [22] were the first to discover that the classical Ford-Fulkerson min-cut/max-flow correspondence [20] could provide global algorithms when the set of labels is binary and the model satisfies certain criteria. The work of [23] 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 [28] (known as the Potts model [31] for the discrete problem). Recently, for this model, Bae and Tai [2] provide a global graph cut method for the case of four labels under some technical assumptions. In this paper, however, our focus is the continuous setting; see [24] 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. [35] 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. [26] used a similar method, with a slightly different regularization and splitting technique, but this had comparable limitations. Most recently, Pock et al. [29] used the dual

formulation of total variation (TV), enforcing a convex constraint on the dual variables. 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, which results in a better approximation. Nevertheless, this method can still not guarantee a globally optimal solution.

The goal of this paper is to provide a convex formulation and an algorithm that is guaranteed to compute a global solution for the multi-phase piecewise constant Mumford-Shah model. To the best of our knowledge, this paper provides the first approach that can make this guarantee.

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. [27] to formulate the problem, which provides a convenient representation to enforce the constraints in the optimization. Second, we use the novel *convexification* approach of Pock et al. [30] (which was motivated by the work of Ishikawa in the discrete Markov Random Field setting [23]) that converts certain non-convex variational problems to equivalent convex problems in a higher dimensional space. Using the dual formulation of TV written in the PCLSM framework allows our method to obtain a convex-concave min-max problem from which a global solution may be computed.

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 describing the representation we use for the multi-phase problem as well as the convexification technique of [30]. Later in the section, we show our main result: that we can find a global solution for multi-phase segmentation; we also describe an algorithm for computing such a solution. Section 4 explains the numerical implementation of the algorithm and shows some experimental results. Finally, in Section 5, we compare our method to previous work in this direction and 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 [28] 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 [31] 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 [33] 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 co-area 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 many of the recent multi-phase approaches mentioned in the introduction; see Section 5 for further discussion.

3 Extension to multi-phase

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 the literature, there are a number of representations of the segmentation problem as such, e.g. [34, 15]. For reasons which we hope to clarify throughout this paper, we use the PCLSM representation of Lie et al [27] in order to globally solve the problem. 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 TV 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 also 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. Our problem is now

$$\min_{u \colon \Omega \to [0,k]} \left\{ \sum_{i=0}^{k} \int_{\Omega} |\nabla \psi_i(u)| + \psi_i(u) g_i(x) \, dx \right\} \quad \text{s.t.} \quad K(u) = 0. \tag{5}$$

Notice there is no harm replacing the condition $u \in \mathbb{R}$ with $u \in [0, k]$ because the constraint $K(u)$ already demands $u \in \{0, \ldots, k\}$. Globally solving this problem appears to be quite formidable since both the objective function and the feasible set are non-convex. The remedy is a technique for converting certain non-convex problems to convex ones, which we describe in the following subsection.

3.2 Convex relaxation by functional lifting

We briefly review the method of [30] to solve minimization problems of the form

$$\min_{u \colon \Omega \to \Gamma := [\gamma_{\min}, \gamma_{\max}]} \int_{\Omega} |\nabla u| + \rho(x, u(x)) \, dx, \tag{6}$$

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. While our problem (5) is not quite of this form, we will eventually use this method in our technique. In our application, we will frequently take $\gamma_{\min} = 0$ and $\gamma_{\max} = k$ since these are the endpoints of the feasible set of (5).

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

$$u(x) = \gamma_{\min} + \int_{\Gamma} \phi(x, \gamma) d\gamma. \tag{7}$$

One tool used to accomplish this reformulation is the co-area formula [18, 19]

$$\int_{\Omega} |\nabla u| dx = \int_{\Omega} \int_{\Gamma} |\nabla \phi| d\gamma dx.$$

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

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

Hence, 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 $u(x) - \gamma$. It follows that

$$\int_{\Omega} \rho(x, u(x))\, dx = \int_{\Omega} \int_{\Gamma} \rho(x, \gamma)\, |\partial_\gamma \phi|\ d\gamma\, dx.$$

Thus, when ϕ is the super-level set function of u, we have

$$\int_{\Omega} |\nabla u| + \rho(x, u(x))\, dx = \int_{\Omega} \int_{\Gamma} |\nabla \phi| + \rho(x, \gamma)\, |\partial_\gamma \phi|\ d\gamma\, dx. \tag{8}$$

This implies that we may solve (6) from

$$\min_{\phi \in \{0,1\}} \int_{\Omega} \int_{\Gamma} |\nabla \phi| + \rho(x, \gamma)\, |\partial_\gamma \phi|\ d\gamma\, dx, \tag{9}$$

as long as we require ϕ to be a super-level set function, i.e., $\phi(x, \gamma_{\min}) = 1, \phi(x, \gamma_{\max}) = 0$, and $\phi(x, \cdot)$ is non-increasing. We then use (7) to obtain u from ϕ. Observe that the objective function of (9) is convex in ϕ because the non-convex function ρ does not depend on ϕ; the non-convexity has somehow been integrated out.

The authors of [30] 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 [23] in which an appropriate auxiliary graph with an extended node set was created, and standard (binary) graph-cut computations (see, e.g., [25]) were used to solve the problem.

Finally, analogous to the procedure in [12], a relaxation on ϕ is taken to allow $\phi \in [0, 1]$. This is crucial to ultimately obtain a convex minimization problem. The form of the integrand of (9) allows for essentially the same proof as that of [12] to show the relaxation is exact, in the sense described at the end of Section 2. In other words, thresholding a minimizer of the convex problem

$$\min_{\phi \in D} \int_\Omega \int_\Gamma |\nabla \phi| + \rho(x, \gamma) \, |\partial_\gamma \phi| \; d\gamma \, dx,$$

where

$$D = \{\phi \colon \Omega \times \Gamma \to [0, 1] \colon \phi(x, \gamma_{\min}) = 1, \phi(x, \gamma_{\max}) = 0\},$$

yields a minimizer of (9), which in turn via (7) results in a solution of the original non-convex problem. It is worth mentioning that because the resulting objective function is convex but not strictly so, there is in general not a unique minimizer.

Despite the effectiveness of this method, it does not directly apply to our segmentation model. The first term in (6) is the TV of the labeling function u. This penalizes boundaries proportional to the difference between the values in the neighboring segments. Instead, our model (as well as many others) wants to penalize all boundaries equally since the value of the labels should be arbitrary.

3.3 Convex method for multi-phase problem

We are now ready to explain our method for finding a global solution of (5). To enforce the constraint $K(u) = 0$, we use an augmented Lagrangian method [3]. Using results from the theory of constrained optimization, if we find a sequence $\{u_j\}$ of global minimizers of

$$\min_{u \in [0,k]} \left\{ \int_\Omega \sum_{i=0}^{k} \left(|\nabla \psi_i(u)| + \psi_i(u) g_i(x) \right) + \lambda_j K(u) + \frac{r_j}{2} K(u)^2 \, dx \right\} \tag{10}$$

for appropriate choices of the *multipliers* $\lambda_j \in L^2(\Omega)$ and the *penalty parameters* $r_j > 0$, then any limit point of the sequence will be a global minimizer of (5). For example, one condition that guarantees such a global minimizer is when the set of multipliers is bounded and the penalty parameters satisfy $r_1 < r_2 < \cdots$ and $r_j \to \infty$ [3, Prop. 2.1]. In practice, appropriately chosen values for λ_j, r_j can provide better convergence results; see Section 4. It remains to find a global minimizer of (10) for fixed j.

To this end, we use the dual formulation of TV [13, 8, 9]. We introduce dual variables $p_i \colon \Omega \to \mathbb{R}^d$ (recall that $\Omega \subset \mathbb{R}^d$) for $i = 0, \ldots, k$ so that (10) becomes

$$\min_{u \in [0,k]} \max_{|p_i| \leq 1} \left\{ \int_\Omega \sum_{i=0}^{k} (-\operatorname{div} p_i \cdot \psi_i(u)) + w_j(x, u(x))\, dx \right\}, \tag{11}$$

where we have put

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

The norm in the maximization constraint in the above expression refers to the L^2 norm on the dual variables.

Using the functional lifting method described in Section 3.2, we see that if ϕ is the super-level set function of u, then

$$\max_{|p_i| \leq 1} \int_\Omega \sum_{i=0}^{k} (-\operatorname{div} p_i \cdot \psi_i(u)) + w_j(x, u(x))\, dx$$

is equal to

$$\max_{|p_i| \leq 1} \left\{ \int_\Omega \int_0^k \left(\sum_{i=0}^{k} (-\operatorname{div} p_i \cdot \psi_i(\gamma)) + w_j(x, \gamma) \right) |\partial_\gamma \phi|\; d\gamma\, dx \right\},$$

and hence (11) is equivalent to

$$\min_{\phi \in D} \max_{|p_i| \leq 1} \left\{ \underbrace{\int_\Omega \int_0^k \left(\sum_{i=0}^{k} (-\operatorname{div} p_i \cdot \psi_i(\gamma)) + w_j(x, \gamma) \right) |\partial_\gamma \phi|\; d\gamma\, dx}_{\Phi_j(\phi, \mathbf{p})} \right\},$$

where

$$D = \{\phi \colon \Omega \times \Gamma \to [0,1] \colon \phi(x,0) = 1, \phi(x,k) = 0\}. \tag{12}$$

Put

$$X = \left\{\mathbf{p} = (p_0, \ldots, p_k) \colon \Omega \to \mathbb{R}^{d\times(k+1)} \colon \ |p_i| \leq 1 \quad \forall i = 0, \ldots, k\right\}. \tag{13}$$

We have an optimization problem over the set $D \times X$, where D and X are compact, convex subsets of linear topological spaces, such that $\Phi_j(\phi, \cdot)$ is concave on X for all $\phi \in D$ and $\Phi_j(\cdot, \mathbf{p})$ is convex on D for all $\mathbf{p} \in X$. By Sion's minimax theorem [32], the minimization and maximization operations may be freely interchanged.

We summarize our main result in the following theorem, whose proof follows from the entirety of this section.

Theorem 1. *The multi-phase segmentation problem*

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

equivalently

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

can be globally solved by finding a limit point u^* of a sequence of functions u_j, constructed as follows. 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 optimization problem

$$\min_{\phi \in D} \max_{\mathbf{p} \in X} \Phi_j(\phi, \mathbf{p})$$

to obtain a function ϕ_j^* and then form the binary function $\phi_j := 1_{\{\phi_j^* > t\}}$ for any $t \in (0,1)$. Finally, compute the solution u_j via

$$u_j(x) = \int_0^k \phi_j(x, \gamma) d\gamma.$$

3.4 Algorithm

Let us now describe an algorithm to obtain a global optimal solution of (1), as guaranteed by Theorem 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.

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

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

or equivalently,

$$\max_{|p_i|\leq 1} \left\{ \int_\Omega p_i \cdot \nabla \left(\int_0^k \psi_i(\gamma) \left|\partial_\gamma \phi\right| \, d\gamma \right) \, dx \right\}.$$

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

$$p_i^{n+1} = \mathcal{P}_X \left(p_i^n + \tau_p \nabla \left(\int_0^k \psi_i(\gamma) \left|\partial_\gamma \phi^n\right| \, d\gamma \right) \right), \tag{14}$$

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 defined in (13), i.e.,

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

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

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

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

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

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

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 $\nabla_\gamma \cdot = \mathrm{div}_\gamma$ 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

$$\phi^{n+1} = \mathcal{P}_D \left\{ \phi^n + \tau_\phi \nabla_\gamma \cdot \left(W(x, \gamma, \mathbf{p}) \frac{\partial_\gamma \phi^n}{\sqrt{\varepsilon^2 + (\partial_\gamma \phi^n)^2}} \right) \right\}, \tag{15}$$

where $\mathcal{P}_D$ is the Euclidean projection onto the convex set D defined in (12). Summarizing, we propose the following algorithm.

Algorithm 1. *1. Fix step sizes $\tau_p, \tau_\phi > 0$, convergence tolerances $0 < \epsilon_u, \epsilon_\phi \ll 1$, and regularization parameter $0 < \varepsilon \ll 1$.*

2. Let j be the iteration number for the function u, multiplier λ, and penalty parameter r. Initialize u^0, λ^0, and r^0. Until convergence $|u^{j+1} - u^j| < \epsilon_u$, update u^j, λ^j, and r^j with the following:

(a) Let n be the iteration number for the variables ϕ and p_i ($i = 0, \ldots, k$). Initialize ϕ^0 as the super-level set function of u^j. Initialize $p_i^0 = 0$. Until convergence $|\phi^{n+1} - \phi^n| < \epsilon_\phi$, update ϕ^n and p_i^n with the following:

i. p_i-step: For each i, update using (14).

ii. ϕ-step: Update using (15).

(b) Threshold ϕ^{n+1} to $1_{\{\phi^{n+1} > \frac{1}{2}\}}$ then obtain u^{j+1} from the layer-cake formula

$$u^{j+1} = \int_0^k 1_{\{\phi^{n+1} > \frac{1}{2}\}} \, d\gamma.$$

Choose λ^{j+1} and r^{j+1} to satisfy the conditions of the theorem.

Note that the Euclidean projection $\mathcal{P}_D$ can be implemented by a simple point-wise truncation operation.

4 Numerical implementation and experimental results

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

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

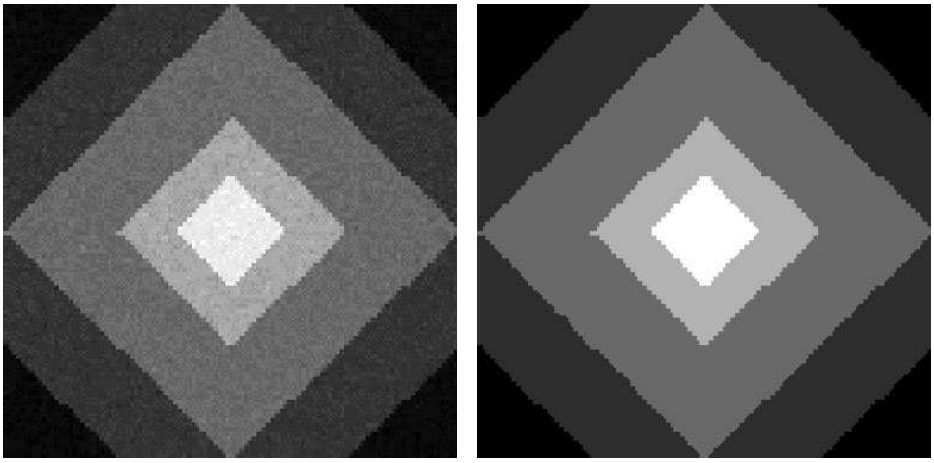
with $N_\gamma = 1 + k/\Delta\gamma$. A spatially continuous function $f \colon \Omega \times \Gamma \to \mathbb{R}$ is now approximated by a discrete function $f^{h,\Delta\gamma}$, from which we will often omit the superscripts. For $f : \{1, \ldots, N\}^2 \times \{1, \ldots, N_\gamma\} \to \mathbb{R}$, 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 [27], the penalty parameters r^j are taken to be equal to some constant r for all j, and the Lagrange multipliers are updated as $\lambda^{j+1} = \lambda^j + rK(u^j)$; we elect to use this update in our implementation. We let $r = 1$ and initialize $\lambda^0 = 1$.

Our first two experiments are given to illustrate that our method can obtain a global solution in situations where previous methods may not. The first is well-known in the discrete graph-cut literature. In [4], an experimental comparison of min-flow/max-cut algorithms is given. In particular, for the



(a) Input (b) Result

Figure 1: Segmentation of the diamond image.

multi-labeling problem of the diamond picture in Figure 1(a), the authors of [4] explain that the problem is NP-hard and instead approximate the solution using the iterative expansion method of [5]. The method is proved to find an approximate solution within a factor of two of the optimal value of the objective function, but cannot find a global solution. On the other hand, our method produces the global minimum. The segmentation result is given in Figure 1(b).

Next, we consider a synthetic three-phase segmentation problem whose optimal solution is given by a triple junction. This example is motivated by an example in [30]. The authors of [30] demonstrated that their method finds a solution that is closer to the optimal solution than the previous techniques of [35, 26]. However, the solution obtained there is not binary (i.e., it does not uniquely assign exactly one label to each point) and hence is not a solution of the original problem. In contrast, our method produces a valid solution by construction and our theory guarantees its optimality. The segmentation result is shown in Figure 2.

We now illustrate our algorithm on two real-world examples. Figure 3(a) shows an MRI brain image common to medical imaging applications. Figure 3(b) shows the result with four phases. We show each phase in a Figures 3(c)–(f). Finally, we apply the algorithm to a color image in Figure 4 using four phases.

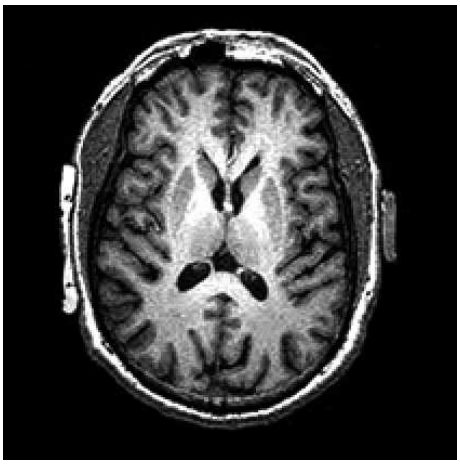
(a) Input (b) Result

Figure 2: Segmentation of a synthetic image. The optimal solution is given by a 120° triple junction.

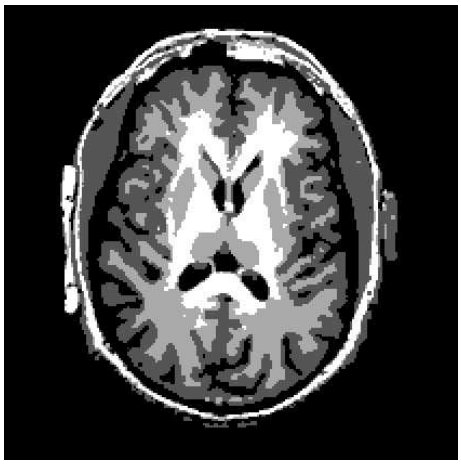
Recall that our method assumes that the region descriptors (which, in the case of piecewise constant Mumford-Shah, are characterized by the mean intensities of each segment) are known *a priori*. In our experiments, we apply a k-means clustering algorithm to determine these values. Variations of our algorithm could include updating these values during the segmentation process, but at the moment our theory would not guarantee an optimal solution. Another alternative would be to run the entire algorithm several times, updating these values after convergence, to yield slightly better results.

5 Comparison with other approaches

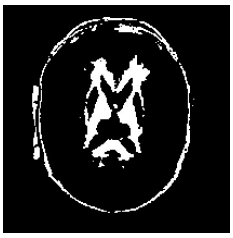
In the previous sections we described a method for globally solving the multi-phase segmentation problem under the Mumford-Shah/Potts model. Our method relied upon a particular way of representing the optimization, namely, the PCLSM representation of Lie et al. [27]. We now contrast our approach with the other ones mentioned in the Introduction, none of which can guarantee a global solution.



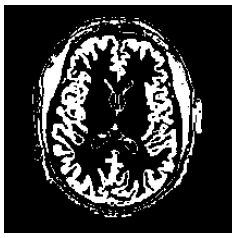
(a) Input



(b) Result

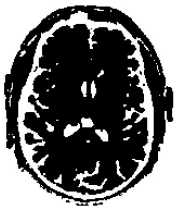


(c) Phase 1



(d) Phase 2

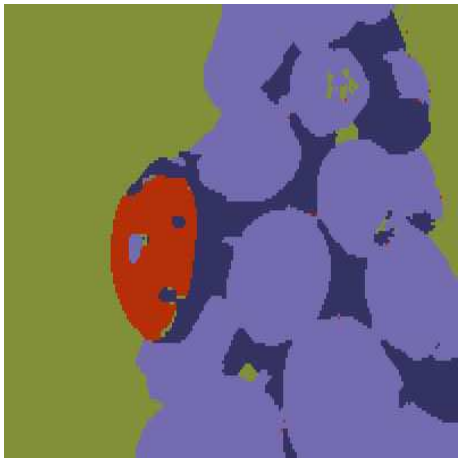
(e) Phase 3



(f) Phase 4

Figure 3: Segmentation of an MRI brain image into four phases.

(a) Input



(b) Result

Figure 4: Segmentation of a natural color image into four phases.

5.1 Convex relaxation over simplex (Zach et al. [35], Lellmann et al. [26])

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 3.3, consider instead the formulation

$$\min_{\boldsymbol{\Psi}=(\psi_0,\dots,\psi_k)\in S} \left\{ F(\Psi) := \sum_{i=0}^{k} \int_\Omega |\nabla \psi_i| + \psi_i g_i(x)\, dx \right\}, \tag{16}$$

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 [30]) 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

$$\min_{(\psi_0,\ldots,\psi_k)\in C} \left\{ \sum_{i=0}^{k} \int_\Omega |\nabla \psi_i(x)| + \psi_i(x) g_i(x)\, dx \right\}. \tag{17}$$

This is essentially the method proposed in [35]. To solve (17), additional variables $(v_0, \ldots, v_k) \colon \Omega \to \mathbb{R}^k$ were introduced to decouple the regularization and data terms [1, 6], yielding

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

where $\mu > 0$ is a parameter that controls the quadratic approximation of ψ_i and v_i for each i. Similarly, Lellmann et al. [26] considered the relaxed problem

$$\min_{(\psi_0,\ldots,\psi_k)\in C} \left\{ \int_\Omega \sqrt{\sum_i |\nabla \psi_i|^2} + \psi_i g_i(x)\, dx \right\}. \tag{18}$$

Up to constant factors, this is equivalent to (17) when $(\psi_0, \ldots, \psi_k) \in S$. To solve the convex optimization problem (18), a Douglas-Rachford splitting algorithm [16] was used.

The issue with both of these methods is that the minimizers of (17) and (18) cannot be guaranteed to lie in the set S. More precisely, if $\Psi^* = (\psi_0^*, \ldots, \psi_k^*)$ is a minimizer of, for example, (17), 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 [30]) to obtain a minimizer of (16) simply by constructing $(1_{\{\psi_0^* > t\}}, \ldots, 1_{\{\psi_k^* > t\}})$ for any $t \in (0, 1)$, since this thresholded vector need not even be in S. It is also not possible to obtain the minimizer by assigning, for each $x \in \Omega$, the unit vector $\vec{\mathbf{e}}_{i*}$ with $i^* = \operatorname{argmin}_j \psi_j(x)$. In short, this relaxation is not exact and therefore cannot be used to globally solve the original problem.

5.2 Convex relaxation on dual variables (Pock et al. [11])

Another decomposition of the labeling function $u\colon \Omega \to \{0, \dots, k\}$ is the following. Put $\theta_0 = 1$, $\theta_{k+1} = 0$, and for $i = 1, \dots, k$,

$$\theta_i(x) = \begin{cases} 1 & \text{if } u(x) \geq i, \\ 0 & \text{otherwise.} \end{cases}$$

Then $u = \sum_i \theta_i$, and a one-to-one correspondence between the labeling function u and the k-tuple $(\theta_1, \dots, \theta_k)$ is guaranteed by imposing the condition $(\theta_1, \dots, \theta_k) \in B_0$, where

$$B_0 = \left\{(\theta_1, \dots, \theta_k)\colon \Omega \to \{0,1\}^k\colon 0 \leq \theta_k(x) \leq \cdots \leq \theta_1(x) \leq 1 \quad \forall x \in \Omega\right\}.$$

Note the relationship between this decomposition and the previous one:

$$\theta_i(x) - \theta_{i+1}(x) = \psi_i(x).$$

Under this framework, the segmentation problem is

$$\min_{(\theta_1,\dots,\theta_k)\in B_0} \left\{ \sum_{i=0}^{k} \int_\Omega |\nabla(\theta_i(x) - \theta_{i+1}(x))| + (\theta_i(x) - \theta_{i+1}(x)) g_i(x)\, dx \right\}. \quad (19)$$

The novel approach of [29] is to use the dual formulation of TV to obtain a more tractable minimization problem from (19). Recall that the regularization term in the Mumford-Shah/Potts model is the $(d-1)$-dimensional Hausdorff measure of the so-called *jump set* J_u (that is, the set of essential discontinuities) of the function u of bounded variation (see [17] for more on BV functions and their properties). In other words, up to constant factors,

$$\sum_{i=0}^{k} \int_\Omega |\nabla(\theta_i(x) - \theta_{i+1}(x))|\, dx = \mathcal{H}^{d-1}(J_u).$$

Although highly nontrivial, this can be written (see the technical report [11])

$$\mathcal{H}^{d-1}(J_u) = \sup_{\xi\in\mathcal{K}} \left\{ \sum_{i=1}^{k} - \int_\Omega \theta_i(x) \operatorname{div} \xi_i(x)\, dx \right\},$$

where

$$\mathcal{K} = \left\{ (\xi_1, \ldots, \xi_k) \colon \Omega \to \mathbb{R}^{d \times k} \colon \left| \sum_{i_1 \leq i \leq i_2} \xi_i(x) \right| \leq 1 \ \forall x \in \Omega,\, 1 \leq i_1 \leq i_2 \leq k \right\}.$$

It is straightforward to see that this set $\mathcal{K}$ is convex. Thus, this yields the problem

$$\min_{\theta \in B_0} \sup_{\xi \in \mathcal{K}} \left\{ \sum_{i=1}^{k} - \int_\Omega \theta_i(x) \operatorname{div} \xi_i(x) + (\theta_i(x) - \theta_{i+1}(x)) g_i(x)\, dx \right\},$$

which turns out to be equivalent (again see [11]) to

$$\min_{\theta \in B} \sup_{\xi \in \mathcal{K}} \left\{ \sum_{i=1}^{k} - \int_\Omega \theta_i(x) \operatorname{div} \xi_i(x) + |\theta_i(x) - \theta_{i+1}(x)|\, g_i(x)\, dx \right\},$$

where

$$B = \left\{ (\theta_1, \ldots, \theta_k) \colon \Omega \to \{0,1\}^k \right\}$$

is binary but now unordered. The problem remains non-convex due to the set B. As we have seen previously, a relaxation is taken to minimize over θ over the convex hypercube

$$R = \left\{ (\theta_1, \ldots, \theta_k) \colon \Omega \to [0,1]^k \right\}.$$

However, once again the relaxation fails to be exact: there is no guarantee that we may obtain a true minimizer over B from a minimizer over R.

5.3 Exact, separately convex relaxation

Finally, let us mention one more alternative approach to use convex relaxation for the multi-phase problem. Suppose the number of phases $k+1$ is equal to 2^n for some integer n. (In fact, this is no restriction whatsoever since superfluous phases may be taken to be empty.) The multi-phase framework [34] uses n level set functions $\varphi_1, \ldots, \varphi_n$ to represent the segmentation, where the union of the zero level-sets of the φ_i yields the edges of the segmentation. Equivalently, we can consider the binary vector of Heaviside functions of the φ_i, say $\mathbf{h} = (h_1, \ldots, h_n) \colon \Omega \to \{0,1\}^n$; each unique binary vector corresponds

to one phase of the partition. The segmentation problem may be formulated as

$$\min_{\mathbf{h}\in\{0,1\}^n} G(\mathbf{h}) := \left\{ \int_\Omega \sum_{i=1}^n |\nabla h_i| + \sum_{\mathbf{b}\in\{0,1\}^n} Z_{\mathbf{b}}(\mathbf{h}) \widetilde{g}_{\mathbf{b}}(x)\, dx \right\}, \tag{20}$$

where

$$Z_{\mathbf{b}}(\mathbf{h}) = \prod_{i=1}^n z_{b_i}(h_i), \qquad z_{b_i}(y) = \begin{cases} 1-y & \text{if } b_i = 0, \\ y & \text{if } b_i = 1. \end{cases}$$

The region descriptor $\widetilde{g}_{\mathbf{b}}$ is simply a relabeling of g_j where the components of $\mathbf{b}$ are the base-2 digits for j. For example, in the four-phase case when $n = 2$,

$$\begin{aligned} G(h_1, h_2) = {} & \int_\Omega |\nabla h_1|\ dx + \int_\Omega |\nabla h_2|\ dx + \int_\Omega g_{(1,1)}(x) h_1(x) h_2(x)\, dx \\ & + \int_\Omega g_{(0,1)}(x)(1 - h_1(x)) h_2(x)\, dx + \int_\Omega g_{(1,0)}(x) h_1(x)(1 - h_2(x))\, dx \\ & + \int_\Omega g_{(0,0)}(x)(1 - h_1(x))(1 - h_2(x))\, dx. \end{aligned}$$

It is straightforward to prove that (20) has an exact convex relaxation.

Proposition 2. *We have*

$$\min_{\mathbf{h}\in\{0,1\}^n} G(\mathbf{h}) = \min_{\mathbf{h}\in[0,1]^n} G(\mathbf{h}).$$

Moreover, if $\mathbf{h}^* \colon \Omega \to [0,1]^n$ *is a solution of the relaxed problem, then*

$$(1_{\{h_1^*(x) > t_1\}}, \ldots, 1_{\{h_n^*(x) > t_n\}}) \colon \Omega \to \{0,1\}^n$$

is a solution of the binary problem (20) for any $(t_1, \ldots, t_n) \in (0,1)^n$.

Proof. We generalize the two-phase proof from [7]. Given a solution $\mathbf{h}^*$ of the relaxed problem, define $\phi(x, \mathbf{t}) = (\phi_1(x, t_1), \ldots, \phi_n(x, t_n))$, where $\phi_i(x, t_i) = 1_{\{h_i^*(x) > t_i\}}$, for any $\mathbf{t} = (t_1, \ldots, t_n) \in [0,1]^n$. By the coarea formula,

$$\int_\Omega |\nabla h_i^*|\ dx = \int_\Omega \int_0^1 |\nabla \phi_i(x,t)|\ dt.$$

By the layer-cake formula,

$$\begin{aligned}\int_\Omega Z_{\mathbf{b}}(\mathbf{h}^*)\widetilde{g}_{\mathbf{b}}(x)\,dx &= \int_\Omega \widetilde{g}_{\mathbf{b}}(x)\left(\prod_{i=1}^{n}\int_0^{z_{b_i}(h_i^*)} dt_i\right)dx \\ &= \int_{[0,1]^n}\int_\Omega \widetilde{g}_{\mathbf{b}}(x) Z_{\mathbf{b}}(\phi(x,\mathbf{t}))\,dx\,d\mathbf{t}.\end{aligned}$$

Hence,

$$G(\mathbf{h}^*) = \int_{[0,1]^n} G(\phi(x,\mathbf{t}))\,d\mathbf{t}.$$

Let $\mathbf{h}$ be a solution of the binary problem. Since $\phi \in \{0,1\}^n$, we have $G(\phi) \geq G(\mathbf{h})$ for all $\mathbf{t}$. On the other hand, because the relaxed problem is minimized over a larger set, we have $G(\mathbf{h}) \geq G(\mathbf{h}^*)$. Thus,

$$G(\mathbf{h}^*) = \int_{[0,1]^n} G(\phi(x,\mathbf{t}))\,d\mathbf{t} \geq \int_{[0,1]^n} G(\mathbf{h}) = G(\mathbf{h}) \geq G(\mathbf{h}^*),$$

which implies that $G(\mathbf{h}) = G(\mathbf{h}^*) = G(\phi(x,\mathbf{t}))$ for all $\mathbf{t} \in (0,1)^n$. □

The objective functional $G(\cdot)$ is not convex, but it is separately convex in each h_i, i.e., for all i, if every other h_j $(j \neq i)$ is fixed, then the function is convex in h_i. Separate convexity cannot in general provide an algorithm that guarantees to find a global minimizer. However, in practice, this method seems to give satisfactory results.

5.4 Conclusion

In conclusion, we have provided a method that guarantees a global solution of the multi-phase piecewise constant Mumford-Shah segmentation problem. The method solves a sequence of convex optimization problems to yield a global solution of the original non-convex problem. Our method appears to be the first in the literature that can guarantee that a global solution is obtained.

In this section, we have described alternative approaches to solving this multi-phase segmentation problem. In some sense, compared with our approach, these methods are more natural multi-phase generalizations of [12]. However, each of these methods is flawed in the sense that a global minimizer cannot be guaranteed. The techniques described in Sections 5.1–5.2 are not

exact relaxations; indeed, in general, one cannot obtain a global minimizer of the multi-phase problem from a minimizer of the respective convex reformulations. On the other hand, the technique described in Section 5.3 is an exact relaxation, but unfortunately the lack of convexity of the reformulation prohibits the assurance of a global minimizer.

We see that the strength of our approach is to formulate the multi-phase problem in such a way that we can apply a convex relaxation approach faithfully. This involves using the PCLSM framework with the augmented Lagrangian, the dual formulation of TV, and functional lifting for the convex relaxation. This combination of tools is able to guarantee the global solution, in contrast to previous methods.

There is still future work that can be done in this direction. This includes developing algorithms to improve the efficiency of our method. We also want to consider the more difficult problem in which the optimal constants (i.e., the region descriptors) are not known *a priori*.

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

[4] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, Sept. 2004.

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, Nov 2001.

[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] G. Chung and L. A. Vese. Energy minimization based segmentation and denoising using a multilayer level set approach. In A. Rangarajan, B. C. Vemuri, and A. L. Yuille, editors, *EMMCVPR*, volume 3757 of *Lecture Notes in Computer Science*, pages 439–455. Springer, 2005.

[16] 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.

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

[18] H. Federer. Curvature measures. *Transactions of the American Mathematical Society*, 93(3):418–491, 1959.

[19] W. H. Fleming and R. Rishel. An integral formula for total gradient variation. *Arch. Math. (Basel)*, 11:218–222, 1960.

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

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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

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

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

[34] 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.

[35] 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.