A Convex Relaxation Method for a Class of Vector-valued Minimization Problems with Applications to Mumford-Shah Segmentation

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

Abstract

The vast majority of problems in image processing and computer vision lead to non-convex minimization problems that possess local minima, thus making it very difficult to obtain a global solution. In this paper, we establish a method for solving a class of non-convex vector-valued optimization problems by reformulating them as convex problems that are equivalent in a way that is made precise. We provide detailed conditions under which the technique introduced in [14] guarantees a solution of the original problem and then apply the general framework to multi-phase image segmentation problems. More precisely, we show that the multi-phase segmentation model of Vese and Chan [32] and the piecewise constant Mumford-Shah model [21] can be formulated as problems of this class. We provide several experimental results to demonstrate that our convex algorithm yields approximate global solutions to these well known non-convex models.

1 Introduction

Energy minimization problems have become increasingly popular for image processing and computer vision applications, including the models of

*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.

Mumford-Shah [21] for image segmentation, Rudin-Osher-Fatemi [29] for image restoration, and Horn-Schunck [15] for image registration. The basic underlying principle is that minimizers of the energy functional correspond to optimal configurations in the particular model. As a result, there has been significant research focusing on minimization techniques for these types of problems. The primary challenge is often a result of the non-convexity of the energy functional, a common property of a number of widely used models.

In general, non-convex functionals are much more difficult to minimize than convex functionals. An important reason is the fact that when a convex function is minimized over a convex set, every locally optimal solution is globally optimal. Moreover, first-order necessary conditions for optimality turn out to be sufficient. As the noted optimizer Rockafellar is often quoted, "the great watershed in optimization isn't between linearity and nonlinearity, but convexity and non-convexity." [28] One of the earliest methods for handling non-convex problems is the technique of LP-relaxation, in which a 0-1 integer program is replaced with the weaker constraint that each variable belong to the interval $[0, 1]$, hence resulting in a linear program whose solution can be used to gain information about the solution to the original integer program.

In the context of image processing, Chan et al. [10] introduced a relaxation method for globally solving the image segmentation problem in which the image is partitioned into two regions, in part based on a result of Strang [30] on continuous maximal flows. For more complicated optimization problems, however, one must first reformulate the problem before invoking a relaxation technique. One such method is to increase the dimension of the problem to remove the non-convexity, an idea often attributed to Ishikawa [16] in the Markov Random Field setting. Pock et al. [24] provided the analogue in the continuous setting and applied the technique to computer vision applications. It turns out this is a particular case of a more general theory, related to the concepts of Cartesian currents [13] and calibration theory [1] in the calculus of variations.

Later, the work of Goldstein et al. [14] attempted to generalize the previous approach of [24] to the case in which the unknown is vector-valued rather than scalar-valued. However, the extension to the vector-valued case is quite non-trivial and the authors could not prove an exact correspondence between the original problem and the embedded higher-dimensional convex formulation. In this work, we state and prove several conditions that are sufficient to guarantee the original non-convex problem can be solved globally. While

the conditions do not hold in the general case, this provides a certificate of optimality that can be quickly verified for a given application.

In this paper, we focus on two particular applications, both of which are multi-phase image segmentation problems, which was our original motivation for studying this class of problems. First, we show that the multi-phase model of Vese and Chan [32] is a direct application of our general framework and demonstrate the effectiveness of our approach. Next, we show that by introducing variables and enforcing constraints via an augmented Lagrangian method, the piecewise constant Mumford-Shah (PCMS) model [21] can be posed as a vector-valued minimization problem in the required form. In this case, it depends on the specific image whether the method produces a global minimizer, which shows a limitation of the general framework. Nevertheless, this provides an important method that adds to the vast literature of convex relaxation approaches [9, 33, 17, 6], none of which can guarantee a global minimizer to PCMS in all cases.

This paper is organized as follows. In Section 2, we describe the class of non-convex problems we will consider and prove conditions under which these problems can be globally solved using a convex formulation. Then, in Section 3, we give the details of an efficient algorithm to solve the convex problem that has well-known convergence results. We apply our method and algorithm to multi-phase image segmentation problems in Section 4, establishing some important new results. Finally, in Section 5, we give some concluding remarks.

2 General Framework

In this section, we describe a method to formulate a class of non-convex optimization problems as equivalent convex optimization problems. Our method closely follows the work of Goldstein et al. [14], which generalizes previous approaches [16, 24] to the case in which the unknown is vector-valued rather than scalar-valued. In short, the idea is to embed the optimization problem into a higher dimensional space. However, significant care is needed to preserve the equivalence between solutions of the embedded problem and of the original problem. Indeed, the main contribution of our work is to provide the technical conditions required for the reformulation and to establish the equivalence rigorously.

The optimization problem we consider is of the form

$$\min_{\overrightarrow{u}} E(\overrightarrow{u}) := \sum_{i=1}^{m} \int_{\Omega} |\nabla u_i| \, dx + \int_{\Omega} \rho(x, \overrightarrow{u}(x)) \, dx, \tag{1}$$

where $\overrightarrow{u} = (u_1, \ldots, u_m) \colon \Omega \to \Gamma := \{0, 1, \ldots, N_1\} \times \cdots \times \{0, 1, \ldots, N_m\}$. In other words, the unknown is a function $\overrightarrow{u} = (u_1, \ldots, u_m)$ defined on a continuous domain $\Omega \subset \mathbb{R}^d$, and each of its components u_i takes values in the discrete set $\{0, 1, \ldots, N_i\}$. In fact, we could take the co-domain of u_i to be any totally ordered finite set and choose consecutive non-negative integer sets beginning at 0 simply for ease of presentation. We assume that the function $\rho \colon \Omega \times \mathbb{R}^m \to \mathbb{R}$ is bounded from below, so that without loss of generality we may assume that ρ is non-negative by adding a constant to E if necessary. However, we make no other assumptions on ρ; in particular, ρ may be non-convex. As was mentioned in Section 1, a wide variety of problems arising in image processing and computer vision can be written in the form (1); see Section 4 for specific applications to image segmentation.

Let us begin to describe the reformulation procedure. To embed (1) into a higher dimensional space, we introduce the function

$$1_{\{\overrightarrow{u}(x) \succeq \overrightarrow{\gamma}\}} := 1_{\{u_1 \geq \gamma_1, \ldots, u_m \geq \gamma_m\}}(x, \overrightarrow{\gamma}) = \begin{cases} 1 & \text{if } u_1 \geq \gamma_1, \ldots, u_m \geq \gamma_m, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

We call such a function a *box function*, since for fixed $x \in \Omega$, the set of points $\overrightarrow{\gamma}$ in the non-negative orthant of $\mathbb{R}^m$ where $1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}}$ is equal to 1 is a hypercube. We call the point $\overrightarrow{u}$ the *principal vertex*, which is on the opposite corner of the hypercube from the origin. This is a multi-dimensional generalization of what is often called a *super-level set function* in the case $m = 1$. It should be clear that there is a one-to-one correspondence between $\overrightarrow{u}$ and its associated box function. In particular, we may use the formula

$$u_i(x) = \sum_{\ell=1}^{N_i} 1_{\{\overrightarrow{u}(x) \succeq \overrightarrow{\gamma}\}}(x, \ell \overrightarrow{e_i})$$

to recover $\overrightarrow{u}$ from the box function, where $\overrightarrow{e_i} \in \mathbb{R}^m$ denotes the ith standard basis vector.

To study properties of box functions more thoroughly, we introduce the set

$$\widetilde{\Gamma} := \{0, \ldots, N_1 + 1\} \times \cdots \times \{0, \ldots, N_m + 1\},$$

which is simply an augmented version of Γ, the co-domain of $\overrightarrow{u}$. We use this set in order to deal with boundary conditions, and we will work with functions ϕ defined on $\Omega \times \widetilde{\Gamma}$. Let the forward difference operators be defined by

$$(D_i\phi)(x, \overrightarrow{\gamma}) = \begin{cases} 0 & \text{if } \gamma_i = N_i + 1, \\ \phi(x, \overrightarrow{\gamma} + \overrightarrow{e_i}) - \phi(x, \overrightarrow{\gamma}) & \text{otherwise.} \end{cases} \tag{3}$$

Finally, we will often work with the set

$$\begin{aligned} C = \Big\{ \phi \colon \Omega \times \widetilde{\Gamma} \to [0,1] \colon \phi(x, \overrightarrow{0}) = 1 \text{ and} \\ \phi(x, \overrightarrow{\gamma}) = 0 \text{ whenever } \gamma_i = N_i + 1 \text{ for some } i \Big\}. \end{aligned} \tag{4}$$

Note that box functions with principal vertex $\overrightarrow{u} \colon \Omega \to \Gamma$ belong to C.

We point out a few properties of box functions in order to rewrite (1) in terms of the box function (2). First, observe that

$$1_{\{u_i \geq \gamma_i\}} = 1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}}(x, \gamma_i \overrightarrow{e_i})$$

for each $i = 1, \ldots, m$. Thus, by a discrete version of the coarea formula [12],

$$\int_\Omega |\nabla u_i| \, dx = \sum_{\ell=1}^{N_i} \int_\Omega \left| \nabla 1_{\{\overrightarrow{u} \succeq \ell \overrightarrow{e_i}\}} \right| \, dx. \tag{5}$$

Next, we examine the result when difference operators are applied to $1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}}$. We see that the mth order mixed difference $D^m_{1,\ldots,m} := D_m \cdots D_1$ maps every $\overrightarrow{\gamma} \in \Gamma$ to 0 except the principal vertex, which gets mapped to $(-1)^m$, i.e.,

$$(D^m_{1,\ldots,m} 1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}})(x, \overrightarrow{\gamma}) = \begin{cases} (-1)^m & \text{if } \overrightarrow{\gamma} = \overrightarrow{u}, \\ 0 & \text{otherwise.} \end{cases}$$

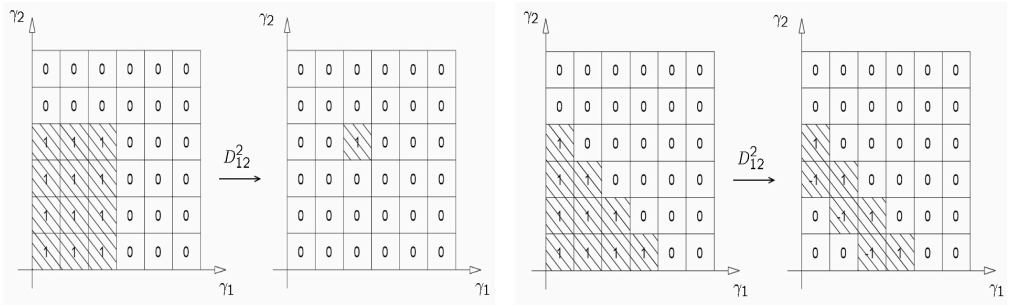
We illustrate this difference operator in Figure 1.

Consequently,

$$(-1)^m \sum_{\overrightarrow{\gamma} \in \Gamma} \rho(x, \overrightarrow{\gamma}) D^m_{1,\ldots,m} 1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}} = \rho(x, \overrightarrow{u}(x)),$$

and so we can rewrite problem (1), using the identities (5) and (2), as

$$\min_{\phi = 1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}}} F(\phi), \tag{6}$$



(a) (b)

Figure 1: An illustration of the difference operator $D^m_{1,\ldots,m}$ when $m = 2$ applied to binary functions. In (a), we see that if ϕ is a box function, then the support of $D^m_{1,\ldots,m}\phi$ is a single point (the principal vertex). On the other hand, in (b) we see that the support of $D^m_{1,\ldots,m}\phi$ will include multiple points for functions ϕ that are not box functions.

where

$$F(\phi) := \sum_{i=1}^{m}\sum_{\ell=1}^{N_i}\int_\Omega |\nabla\phi(x,\ell\overrightarrow{e_i})|\,dx + (-1)^m \sum_{\overrightarrow{\gamma}\in\Gamma}\int_\Omega \rho(x,\overrightarrow{\gamma})D^m_{1,\ldots,m}\phi\,dx. \quad (7)$$

That is, (1) is equivalent to an optimization problem over box functions defined on a space with an additional m dimensions. Moreover, while the original objective function was possibly non-convex in $\overrightarrow{u}$ (due to the function ρ), the reformulated objective function is convex in $\phi = 1_{\{\overrightarrow{u}\succeq\overrightarrow{\gamma}\}}$. However, the minimization is conducted over the non-convex set of box functions. In order to obtain a convex minimization problem, we must change the set over which the minimization is conducted to a set which is convex.

The procedure of allowing the optimization to be taken over a larger set is known as *relaxation*. In general, relaxation introduces minimizers that lie outside the original constraint set and that have no relationship to minimizers of the original minimization problem. Under some special circumstances, a solution of the original problem may be obtained precisely from a solution of the relaxed problem. In this case, the relaxation is said to be *exact*.

Our goal is to find the conditions under which we have an exact relaxation of our problem. To this end, we will establish a series of several preliminary results. The first result is that the functional F defined in (7) satisfies a

generalized coarea formula of the form

$$F(\phi) = \int_0^1 F(1_{\{\phi \geq t\}})\, dt. \tag{8}$$

Let us remark that most versions of the coarea formula use strict thresholding and instead write

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

The reason for our unconventionality is in view of the other results throughout this section.

Lemma 1. *Let F be defined by (7). Then F satisfies the generalized coarea formula (8).*

Proof. Recall that the coarea formula [12] for functions g of bounded variation states that

$$\int_\Omega |\nabla g|\, dx = \int_{-\infty}^{\infty} \int_\Omega \left|\nabla 1_{\{g>t\}}\right|\, dx\, dt.$$

Using some elementary facts about total variation, namely that

$$\int_\Omega |\nabla g|\, dx = \int_\Omega |\nabla(-g)|\, dx = \int_\Omega |\nabla(g+c)|\, dx$$

for any constant $c \in \mathbb{R}$, we see that

$$\begin{aligned}\int_\Omega |\nabla g|\, dx &= \int_{-\infty}^{\infty} \int_\Omega \left|\nabla 1_{\{-g>t\}}\right|\, dx\, dt = \int_{-\infty}^{\infty} \int_\Omega \left|\nabla 1_{\{g<t\}}\right|\, dx\, dt \\ &= \int_{-\infty}^{\infty} \int_\Omega \left|\nabla(1 - 1_{\{g\geq t\}})\right|\, dx\, dt = \int_{-\infty}^{\infty} \int_\Omega \left|\nabla 1_{\{g\geq t\}}\right|\, dx\, dt.\end{aligned}$$

Since $\phi \in [0,1]$, we obtain

$$\sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \int_\Omega |\nabla \phi(x, \ell \overrightarrow{e_i})|\, dx = \int_0^1 \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \int_\Omega \left|\nabla 1_{\{\phi(x,\ell\overrightarrow{e_i})\geq t\}}\right|\, dx\, dt.$$

We also have $\phi(x, \overrightarrow{\gamma}) = \int_0^{\phi(x,\overrightarrow{\gamma})}\, dt = \int_0^1 1_{\{\phi\geq t\}}\, dt$, often referred to as the layer-cake formula [19, p.26-27]. Thus, by linearity,

$$\begin{aligned}(-1)^m &\sum_{\overrightarrow{\gamma}\in\Gamma} \int_\Omega \rho(x, \overrightarrow{\gamma}) D^m_{1,\ldots,m} \phi(x, \overrightarrow{\gamma})\, dx \\ &= \int_0^1 (-1)^m \sum_{\overrightarrow{\gamma}\in\Gamma} \int_\Omega \rho(x, \overrightarrow{\gamma}) D^m_{1,\ldots,m} 1_{\{\phi\geq t\}}\, dx\, dt,\end{aligned}$$

and (8) follows. □

The use of coarea formulas is a ubiquitous tool in convex relaxation methods found in the literature because it can be used to prove properties about minimizers of the functional.

Lemma 2. *Let Y be any subset of functions $\phi\colon \Omega \times \widetilde{\Gamma} \to [0,1]$ and let ϕ^* be any minimizer of F over Y. If $1_{\{\phi \geq t\}} \in Y$ for all $t \in [0,1]$, then $F(\phi^*) = F(1_{\{\phi \geq t\}})$ for all $t \in (0,1]$ and thus $1_{\{\phi \geq t\}}$ is a minimizer of F over Y for all $t \in (0,1]$.*

Proof. Let ϕ^* be a minimizer of F over Y and let ϕ' be a minimizer of F over $Y' = Y \cap \{\phi \in \{0,1\}\}$. Since $Y' \subset Y$, we have $F(\phi^*) \leq F(\overline{\phi})$. By minimality, $F(\phi^*) \leq F(1_{\{\phi^* \geq t\}})$, and so from (8) it follows that

$$F(\phi^*) = \int_0^1 F(1_{\{\phi^* \geq t\}})\, dt \geq \int_0^1 F(\phi')\, dt = F(\phi') \geq F(\phi^*).$$

Hence, all inequalities in the above expression are equalities, and $1_{\{\phi^* \geq t\}}$ is a minimizer of F for almost every $t \in [0,1]$. But for all $t \in (0,1]$ there exists a strictly increasing sequence $\{t_i\}_{i=1}^{\infty}$ converging to t such that $1_{\{\phi^* \geq t_n\}}$ is a minimizer of F for all n and $1_{\{\phi^* \geq t_n\}}$ converges to $1_{\{\phi^* \geq t\}}$ almost everywhere as $n \to \infty$. Invoking lower semi-continuity of total variation and Lebesgue's dominated convergence theorem, we may conclude that $1_{\{\phi^* \geq t\}}$ is a minimizer of F for all $t \in (0,1]$. □

We would like to apply the previous lemma to a convex set Y so that $\min_{\phi \in Y} F(\phi)$ is a convex minimization problem. We would also want binary sets in Y to be box functions to preserve the correspondence to (6), and hence to (1). Recall the difference operator property of box functions, namely, that $D^m_{1,\ldots,m} 1_{\{\vec{u} \geq \vec{\gamma}\}}$ vanishes everywhere except at the principal vertex, where it equals $(-1)^m$. It is thus natural to consider the set

$$X = \{\phi \in C\colon (-1)^m D^m_{1,\ldots,m} \phi \geq 0\}. \tag{9}$$

In particular, if ϕ is binary, then $\phi \in X$ if and only if ϕ is a box function. This leads to the following proposition.

Proposition 3. *Let ϕ^* be a minimizer of $\min_{\phi \in X} F(\phi)$. If $1_{\{\phi^* \geq t\}}$ is a box function for all $t \in (0,1]$, then $1_{\{\phi^* \geq t\}}$ is a minimizer of (6) for all $t \in (0,1]$.*

There are many shortcomings of this proposition. In general, there is no guarantee that all of the thresholded functions will be box functions unless the minimizer ϕ^* happens to be binary. Otherwise, verifying the hypothesis involves checking infinitely many conditions. Alternatively, we can construct a minimizer of (6) without appealing to the coarea formula (5).

Proposition 4. *Let ϕ^* be a minimizer of $\min_{\phi\in X} F(\phi)$. Suppose that there exists $t \in [0,1]$ such that $1_{\{\phi^*\geq t\}}$ is a box function and $F(\phi^*) = F(1_{\{\phi^*\geq t\}})$. Then $1_{\{\phi^*\geq t\}}$ is a minimizer of (6).*

The reason this proposition is useful is that when $\phi \in X$, we can prove there exists $t \in (0,1]$ such that $1_{\{\phi\geq t\}}$ is a box function (although it should be noted that $1_{\{\phi\geq t\}}$ is not necessarily a box function for all t). In fact, we will show that $1_{\{\phi\geq 1\}}$ is a box function.

Lemma 5. *For all $x \in \Omega$, suppose that $\phi\colon \Omega \times \widetilde{\Gamma} \to [0,1]$ lies in the set X defined in (9), i.e. ϕ satisfies the boundary conditions*

(a) $\phi(x, \overrightarrow{0}) = 1$,

(b) $\phi(x, \gamma_1, \ldots, \gamma_m) = 0$ *whenever* $\gamma_i = N_i + 1$ *for some* i,

and ϕ satisfies the difference condition

$$(-1)^m D^m_{1,\ldots,m}\phi \geq 0. \tag{10}$$

Then $1_{\{\phi\geq 1\}}$ is a box function, i.e., there exists a unique $\overrightarrow{u}(x)$ such that $1_{\{\phi\geq 1\}} = 1_{\{\overrightarrow{u}(x)\succeq\overrightarrow{\gamma}\}}$.

Proof. We split up the proof into two parts. First, we show that (10) implies an analogous condition for all differences of all lower orders. Then, we use the conditions on all first and second order differences to prove the desired result.

The first claim is that the difference condition (10) implies

$$(-1)^k D^k_S\phi \geq 0 \text{ for all } 1 \leq k \leq m \text{ and } S \subset \{1,\ldots,m\} \text{ with } |S| = k. \tag{11}$$

Inductively, and by relabeling indices if necessary, it suffices to show that (10) implies $(-1)^{m-1} D^{m-1}_{1,\ldots,m-1}\phi \geq 0$. By way of contradiction, we suppose

that there exists $x \in \Omega$ and $\overrightarrow{\gamma} \in \Gamma$ such that $(-1)^{m-1} D^{m-1}_{1,\ldots,m-1}\phi(x, \overrightarrow{\gamma}) < 0$. We see that

$$\begin{aligned} 0 &\geq (-1)^{m-1} D^{m}_{1,\ldots,m}\phi(x, \overrightarrow{\gamma}) \\ &= (-1)^{m-1} \left(D^{m-1}_{1,\ldots,m-1}\phi(x, \overrightarrow{\gamma} + \overrightarrow{e_m}) - D^{m-1}_{1,\ldots,m-1}\phi(x, \overrightarrow{\gamma}) \right) \\ &> (-1)^{m-1} D^{m-1}_{1,\ldots,m-1}\phi(x, \overrightarrow{\gamma} + \overrightarrow{e_m}). \end{aligned}$$

Applying this recursively, we conclude that there exists $\overrightarrow{\gamma}' = (\gamma'_1, \ldots, \gamma'_m)$ with $\gamma'_m = N_m + 1$ such that $(-1)^{m-1} D^{m-1}_{1,\ldots,m-1}\phi(x, \overrightarrow{\gamma}') < 0$. However, from the boundary conditions we must have $D^{m-1}_{1,\ldots,m-1}\phi(x, \overrightarrow{\gamma}') = 0$, a contradiction.

Our previous work has shown that, in particular,

(c) $D_i\phi \leq 0$ for all $1 \leq i \leq m$, and

(d) $D^2_{ij}\phi \geq 0$ for all $i \neq j$.

We now use (c) and (d) to prove the lemma. To help explain the rest of the proof, we illustrate the remaining steps in Figure 2.

Recall the boundary conditions required by (a) and (b); see Figure 2(a). For each $x \in \Omega$, let $\phi_x(\overrightarrow{\gamma}) = \phi(x, \overrightarrow{\gamma})$. Observe that for all $1 \leq i \leq m$, there exists n_i, depending on x, such that $\phi_x(n_i \overrightarrow{e_i}) = 1$, $\phi_x(j \overrightarrow{e_i}) = 1$ for all $0 \leq j < n_i$, and $\phi_x(j \overrightarrow{e_i}) < 1$ for all $n_i < j \leq N_i + 1$. Indeed, this is an immediate consequence of (a), which requires $\phi_x(\overrightarrow{0}) = 1$, and (c), which requires the list of values

$$\phi_x(\overrightarrow{0}), \phi_x(\overrightarrow{e_i}), \ldots, \phi_x((N_i + 1)\overrightarrow{e_i})$$

to be non-increasing. See Figure 2(b).

Next, we claim that $1_{\{\phi \geq 1\}}$ is a box function with principal vertex $\overrightarrow{n} := (n_1, \ldots, n_m)$. The second order property (d) yields the implication

$$\phi_x(\overrightarrow{\gamma} + \overrightarrow{e_i}) = 1 \text{ and } \phi_x(\overrightarrow{\gamma} + \overrightarrow{e_j}) = 1 \quad \Longrightarrow \quad \phi_x(\overrightarrow{\gamma} + \overrightarrow{e_i} + \overrightarrow{e}_j) = 1 \quad (12)$$

for all $\overrightarrow{\gamma}$ and indices $i \neq j$ in which these expressions are defined. Note that the left-hand side implicitly means that $\phi_x(\overrightarrow{\gamma}) = 1$, due to (a). One can visualize the (i, j)-plane of the values of ϕ_x as a matrix whose first column and last row are equal to 1 from the first observation of this proof. Property (12) forces the remaining interior entries to be subsequently filled with 1's until we form the "box" of 1's. See Figure 2(c).

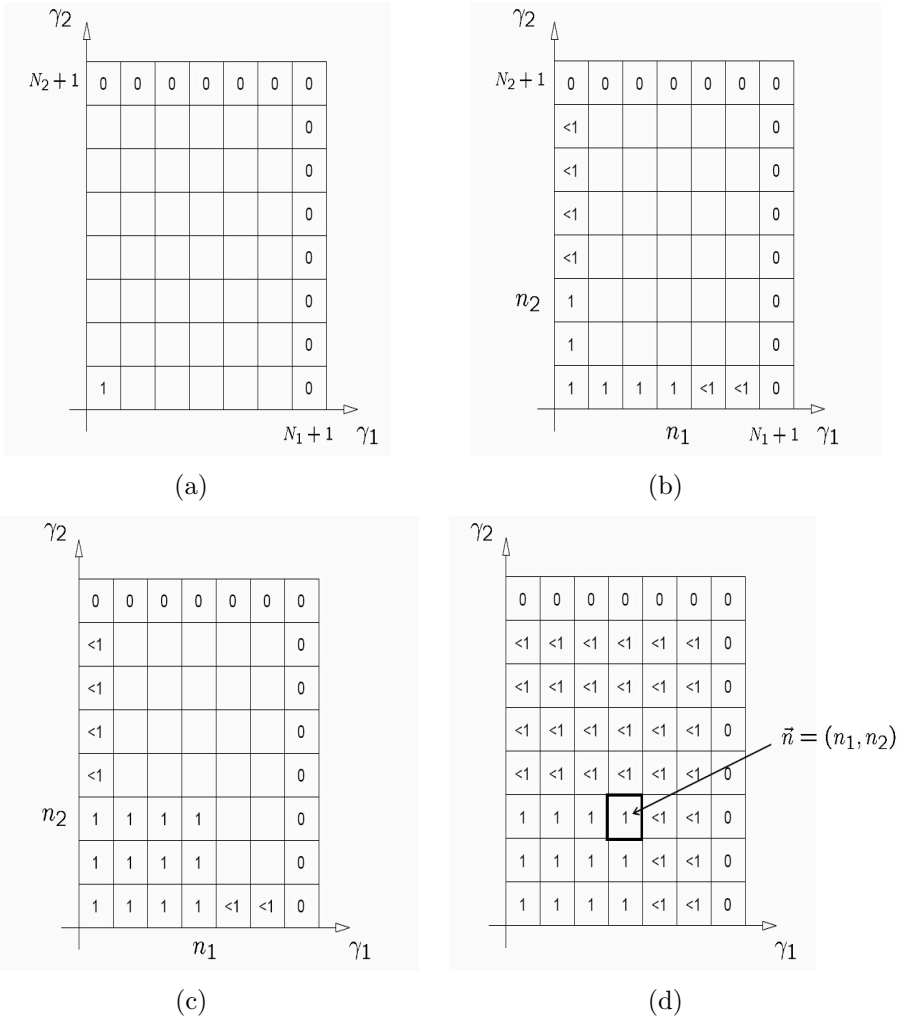
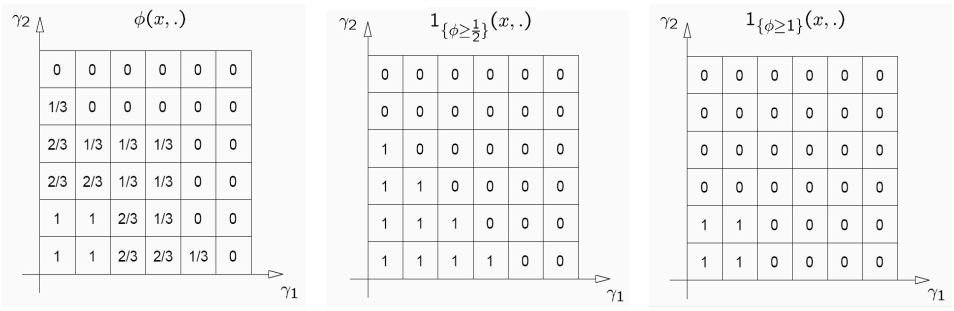


Figure 2: An illustration of the proof of Lemma 5. See the proof for the explanations of each of the steps (a)-(d). In particular, (d) shows $\vec{n}$, which we see is the principal vertex of $1_{\{\phi \geq 1\}}$.



(a) (b) (c)

Figure 3: The function $\phi(x,\cdot)$ in (a) satisfies the hypotheses of Lemma 5. However, in (b) we see that $1_{\{\phi\geq\frac{1}{2}\}}$ is not a box function. As shown in (c), $1_{\{\phi\geq1\}}$ is a box function, as guaranteed by the lemma.

More formally, consider the following contradiction argument. Suppose there exists $\overrightarrow{\gamma} \preceq \overrightarrow{n}$ such that $\phi_x(\overrightarrow{\gamma}) < 1$. Applying the contrapositive of (12), we see that either $\phi_x(\overrightarrow{\gamma} - \overrightarrow{e_i}) < 1$ or $\phi_x(\overrightarrow{\gamma} - \overrightarrow{e_j}) < 1$. This procedure can be used successively to all pairs of distinct indices until we eventually have $\phi_x(\overrightarrow{\alpha}) < 1$ and $\overrightarrow{\alpha}$ equal to $\alpha_i \overrightarrow{e_i}$ for some i with $\alpha_i \leq n_i$. This violates our first observation. Consequently, we have now established that $\phi_x(\overrightarrow{\gamma}) = 1$ for all $\overrightarrow{\gamma} \preceq \overrightarrow{n}$.

Finally, we need only show that if $\overrightarrow{\gamma}$ is such that $\gamma_i > n_i$ for some i (i.e., $\overrightarrow{\gamma} \not\preceq \overrightarrow{n}$), then necessarily $\phi_x(\overrightarrow{\gamma}) < 1$. But this follows quickly from (a) once more, since otherwise this would force $\phi_x(\gamma_i \overrightarrow{e_i}) = 1$, which conflicts with the construction of n_i. See Figure 2(d). We conclude that $1_{\{\phi\geq1\}}$ is a box function with principal vertex $\overrightarrow{n}$, and the proof is complete. □

As we noted before this lemma, it is not necessarily the case that $1_{\{\phi\geq t\}}$ is a box function for all $t \in (0,1]$, even when ϕ satisfies our assumptions. Indeed, consider the example shown in Figure 3. We see that ϕ satisfies the hypotheses of Lemma 5, but that, for example, $1_{\{\phi\geq\frac{1}{2}\}}$ is not a box function. However, as guaranteed by the lemma, $1_{\{\phi\geq1\}}$ is a box function. In contrast, when $m = 1$, the assumption on ϕ is simply that it is non-increasing in γ, and it is immediately evident this is sufficient to guarantee $1_{\{\phi\geq t\}}$ is a box function for all t (cf. [14, Lemma 1]). In fact, it turns out that when $m = 1$, minimizers of F are automatically non-increasing in γ, since it can be shown that $F(\Pi\phi) \leq F(\phi)$ where Π is projection onto the set where $D\phi \geq 0$ (cf. [9, Proposition 4.3]). This is not the case for $m > 1$ and illustrates further

that there are many subtleties to consider when extending the theory to the vector-valued setting.

Combining Proposition 4 and Lemma 5, we can now state our main result.

Theorem 6. *Let ϕ^* be a minimizer of F over X, the convex set of functions $\phi\colon \Omega \times \widetilde{\Gamma} \to [0,1]$ such that*

$$\phi(x, \overrightarrow{0}) = 1 \text{ and } \phi(x, \gamma_1, \ldots, \gamma_m) = 0 \text{ whenever } \gamma_i = N_i + 1 \text{ for some } i$$

and

$$(-1)^m D^m_{1,\ldots,m}\phi \geq 0.$$

Then

$$u_i(x) = \sum_{\ell=1}^{N_i} 1_{\{\phi^* \geq 1\}}(x, \ell\overrightarrow{e_i})$$

is a minimizer of (1) provided $F(\phi^*) = F(1_{\{\phi^* \geq 1\}})$.

We consider this theorem the main result of this section because it provides us with an optimality *certificate*. We may solve a convex minimization problem and perform one straightforward computation to check whether we can obtain a global solution to the original vector-valued non-convex minimization problem (1). We will see in Section 4 that for many applications the certificate condition is nearly satisfied, leading to approximate solutions to non-convex minimization problems that are very close to optimal and independent of the initial condition.

When $m = 1$, the results of this section reduce to the method proposed in Pock et al. [24]. In that case, we need not verify the certificate condition of the theorem. Instead, we observe that $1_{\{\phi \geq t\}} \in X$ whenever $\phi \in X$ for all $t \in (0,1]$, which means we may simply apply Proposition 3 since we know the hypotheses are satisfied. However, when $m > 1$, this is no longer the case.

Let us reiterate that the boundary conditions demanded by the set X is exactly the set C defined in (4). These conditions are quite natural, since they are satisfied by box functions, and they are invariant under thresholding (i.e., if ϕ satisfies the boundary conditions, so does $1_{\{\phi \geq t\}}$ for all t).

Lastly, consider the ramifications of the implication (10) $\Rightarrow$ (11). In particular, this means that instead of minimizing over X, we could obtain the same result by minimizing over $X' \subsetneq X$, the set of ϕ such that $D_i\phi \leq 0$ for all i and $D^2_{i,j}\phi \geq 0$ for all $i \neq j$ with the appropriate boundary conditions.

In fact, we can minimize over any convex set containing X'. However, the single constraint $(-1)^m D^m_{1,\dots,m}\phi \geq 0$ is elegant as well as practical from a computational point of view; we have just one difference constraint to enforce, rather than, say, the $m(m-1)/2$ second order constraints.

3 Algorithm

In this section, we describe an algorithm to solve the discrete version of our convex problem

$$\min_{\phi \in X} F(\phi), \tag{13}$$

where F is given by (7). We write (13) as a saddle point problem

$$\min_{\phi \in C} \max_{|\overrightarrow{p_i}| \leq 1, (-1)^m p_\gamma \leq \rho} \left\{ \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \int_\Omega \overrightarrow{p_i} \cdot \nabla \phi(x, \ell \overrightarrow{e_i})\, dx + (-1)^m \sum_{\overrightarrow{\gamma} \in \Gamma} p_\gamma D^m \phi \right\} \tag{14}$$

and solve it using a primal-dual algorithm. Recall the set C was defined in (4). Since the difference condition required by the set X is enforced by the dual variable p_γ, we need only enforce the boundary conditions on the primal variable ϕ.

To see (14), we may use the Legendre-Fenchel transformation [27]. Consider the function

$$f(z) = \begin{cases} \infty & (-1)^m z < 0, \\ (-1)^m z \rho & (-1)^m z \geq 0. \end{cases}$$

We see that the Legendre-Fenchel transform (or convex conjugate), $f^*(p)$, is equal to

$$\max_z \{pz - f(z)\} = \max_{(-1)^m z \geq 0} \{pz - (-1)^m z\rho\} = \max_{(-1)^m z \geq 0} \{z(p - (-1)^m \rho)\}.$$

If m is even, we have

$$f^*(p) = \max_{z \geq 0} \{z(p-\rho)\} = \begin{cases} 0 & p \leq \rho, \\ \infty & else, \end{cases}$$

and if m is odd,

$$f^*(p) = \max_{z \le 0}\{z(p+\rho)\} = \begin{cases} 0 & p \ge -\rho, \\ \infty & else. \end{cases}$$

In either case,

$$f^*(p) = \begin{cases} 0 & (-1)^m p \le \rho, \\ \infty & \text{else.} \end{cases}$$

Since F is convex, we have

$$f(z) = f^{**}(z) = \max_p\{zp - f^*(p)\} = \max_{(-1)^m p \le \rho} zp.$$

We apply this to the case where $z = D^m_{1,\dots,m}$. For the other terms of the functional, we may either use the dual formulation of total variation [8] or the Legendre-Fenchel transform applied to $f(z) = |z|$.

We discretize the problem in the usual way and consider for simplicity the case $d = 2$ and $\Omega = [0,1]^2$. Let $\Omega^h = \{0,\dots,N_x\} \times \{0,\dots,N_y\}$. The discrete version of (14) is thus

$$\min_{\phi^h \in C^h} \max_{\overrightarrow{p} \in D^h} \left\{ \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \sum_{x^h \in \Omega^h} \overrightarrow{p_i} \cdot \nabla^h \phi^h(x^h, \ell \overrightarrow{e_i}) + (-1)^m \sum_{\overrightarrow{\gamma} \in \Gamma} \sum_{x^h \in \Omega^h} p_\gamma D^m_{1,\dots,m} \phi^h \right\},$$

where ∇^h is the discrete gradient operator,

$$C^h = \Big\{ \phi^h \colon \Omega^h \times \widetilde{\Gamma} \to [0,1] \colon \phi^h(x^h, \overrightarrow{0}) = 1 \text{ and } \phi^h(x^h, \overrightarrow{\gamma}) \text{ whenever } \gamma_i = N_i + 1 \text{ for some } i \Big\},$$

and

$$D^h = \Big\{ \overrightarrow{p} = (\overrightarrow{p_1}, \dots, \overrightarrow{p_m}, p_\gamma) \colon \Omega^h \times \widetilde{\Gamma} \to \mathbb{R}^2 \times \cdots \times \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^{2m+1} \colon |\overrightarrow{p_i}| \le 1, (-1)^m p_\gamma \le \rho \text{ and } \overrightarrow{p_i}(x^h, \overrightarrow{\gamma}) = 0 \text{ if } \gamma_i \ne 0 \Big\}.$$

This can be written

$$\min_{\phi^h \in C^h} \max_{\overrightarrow{p} \in D^h} \langle A\phi^h, \overrightarrow{p} \rangle_{\mathbb{R}^{2m+1}}, \tag{15}$$

where A is the linear operator that maps

$$\phi^h \mapsto (\nabla^h \phi^h, \dots, \nabla^h \phi^h, D^m_{1,\dots,m} \phi^h) \in \mathbb{R}^{2m+1}.$$

Following [23, 31], we use the following modified Arrow-Hurwicz-Uzawa algorithm [25, 2]; see [11] for an excellent overview of primal-dual algorithms. Choose time steps $\tau_\phi, \tau_p > 0$ such that $\tau_\phi \tau_p \|A\|^2 < 1$, where

$$\|A\| = \sup_{\phi^h \neq 0} \frac{\left\|A\phi^h\right\|}{\|\phi^h\|}$$

is the operator norm of A. Along the same lines as [31, Theorem 4.1], we see that

$$\|A\|^2 = \frac{4}{h_x^2} + \frac{4}{h_y^2} + \frac{4^m}{\prod_{i=1}^m h_{\gamma_i}^2},$$

where h_x and h_y are the spatial step sizes of Ω^h and $h_{\gamma_1}, \ldots, h_{\gamma_m}$ are the step sizes of $\widetilde{\Gamma}$. In our case, $h_x = 1/N_x$, $h_y = 1/N_y$, and $h_{\gamma_i} = 1$, and the time step requirement reduces to

$$\tau_\phi \tau_p < \frac{1}{4N_x^2 + 4N_y^2 + 4^m}.$$

Choose any initial values $((\phi^h)^0, (\overrightarrow{p})^0) \in C^h \times D^h$ and put $(\overline{\phi}^h)^0 = (\phi^h)^0$. Then for $n > 0$ use the update scheme

$$\begin{cases} (\overrightarrow{p})^{n+1} = \Pi_{D^h}\left((\overrightarrow{p})^n + \tau_p A(\overline{\phi}^h)^n\right) \\ (\phi^h)^{n+1} = \Pi_{C^h}\left((\phi^h)^n - \tau_\phi A^*(\overrightarrow{p})^{n+1}\right) \\ (\overline{\phi}^h)^{n+1} = 2(\phi^h)^{n+1} - (\phi^h)^n \end{cases},$$

where A^* is the adjoint of A. The operators Π_{D^h} and Π_{C^h} are the projections onto the convex sets D^h and C^h, respectively. Explicitly,

$$\Pi_{D^h}(\overrightarrow{p}) = \left(\frac{p_1}{\max(|(p_1, \ldots, p_m)|, 1)}, \ldots, \frac{p_1}{\max(|(p_1, \ldots, p_m)|, 1)}, \widehat{p_\gamma}\right),$$

with

$$\widehat{p_\gamma} = \begin{cases} \max(p_\gamma, -\rho) & \text{if } m \text{ odd}, \\ \min(p_\gamma, \rho) & \text{if } m \text{ even}. \end{cases}$$

The projection $\Pi_{C^h}(\phi^h)$ is a simple truncation of ϕ^h to the interval [0,1] and setting the boundary conditions $\phi^h(x, \overrightarrow{0}) = 1$ and $\phi^h(x, \overrightarrow{\gamma}) = 0$ if $\gamma_i = N_i + 1$ for some i. As $n \to \infty$, this scheme is guaranteed to converge to a solution of (15) (cf. [23, Theorem 4.1]).

4 Applications to Image Segmentation

In this section, we apply the general framework and algorithm developed in the previous sections to multi-phase image segmentation problems. In particular, our focus will be on non-convex energy minimization problems based on the Mumford-Shah model [21]

$$\inf_{f,K} \int_\Omega (I(x) - f(x))^2 \, dx + \nu \int_{\Omega \setminus K} |\nabla f(x)|^2 \, dx + \mu |K| \tag{16}$$

to find an optimal piecewise smooth approximation of an image $I \colon \Omega \to \mathbb{R}$, where ν and μ are fixed parameters. The edge set $K \subset \Omega$ is a closed set that defines a partition $\Omega = \cup_i \Omega_i$ such that the restrictions of the function f to the phases Ω_i are differentiable. If the function f is taken to be constant on each phase, this reduces to the piecewise constant Mumford-Shah problem

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

where $f = \sum_i c_i 1_{\Omega_i}$. We will assume that the optimal constants c_i are known *a priori* and the number of segments (say, n) is fixed. Even making this assumption leaves us with the difficult non-convex problem

$$\inf_{\Omega_0,\ldots,\Omega_{n-1}} \left\{ \sum_{i=0}^{n-1} \frac{\mu}{2} \, |\partial \Omega_i| + \int_{\Omega_i} (c_i - I(x))^2 \, dx \right\}, \tag{17}$$

where it is implicit that $\Omega = \cup_i \Omega_i$ and the Ω_i are pairwise disjoint. This is known as the Potts model [26] in the discrete setting and has been shown to be an NP-hard problem [5].

We first consider the Vese-Chan multi-phase model [32], which we will see is a direct application of the framework established in the previous sections.

4.1 Vese-Chan multi-phase segmentation

In [32], Vese and Chan proposed a level-set method [22] to solve the piecewise constant problem (17). They suppose $n = 2^m$ and introduce functions $\varphi_1, \ldots, \varphi_m \colon \Omega \to \mathbb{R}$. The boundaries of the phases are represented by the zero-level sets, i.e.,

$$\bigcup_{i=0}^{n-1} \partial \Omega_i = \bigcup_{k=1}^{m} \{x \in \Omega \colon \varphi_k(x) = 0\}.$$

Each $x \in \Omega$ belongs to the phase Ω_i if and only if $i = \sum_{k=1}^{m} 2^k H(\varphi_k(x))$, where H is the Heaviside function

$$H(z) = \begin{cases} 1 & \text{if } z \geq 0, \\ 0 & \text{else.} \end{cases}$$

Let $\overrightarrow{b}(i) = (b_1, \ldots, b_m)$ be the binary representation of $i \in \{0, \ldots, n-1\}$. Let $\omega_0(z) = z$ and $\omega_1(z) = 1 - z$. Then the Vese-Chan minimization problem is

$$\min_{\varphi_1,\ldots,\varphi_m \colon \Omega \to \mathbb{R}} \sum_{k=1}^{m} \int_\Omega |\nabla H(\varphi_k)| \, dx + \sum_{i=0}^{n-1} \int_\Omega \prod_{k=1}^{m} \omega_{b_k(i)}(H(\varphi_k)) \, dx.$$

Since the functional depends only on $H(\varphi_k)$ for each k, this can be rewritten

$$\min_{u_1,\ldots,u_m \colon \Omega \to \{0,1\}} \sum_{k=1}^{m} \int_\Omega |\nabla u_k| \, dx + \sum_{i=0}^{n-1} \int_\Omega \prod_{k=1}^{m} \omega_{b_k(i)}(u_k) \, dx. \tag{18}$$

For example, when $m = 2$ (four phases), we have

$$\begin{aligned} \min_{u_1,u_2 \colon \Omega \to \{0,1\}} \Bigg\{ & \int_\Omega |\nabla u_1| \, dx + \int_\Omega |\nabla u_2| \, dx + \int_\Omega (c_0 - I(x))^2 u_1(x) u_2(x) \, dx \\ & + \int_\Omega (c_1 - I(x))^2 u_1(x)(1 - u_2(x)) \, dx + \int_\Omega (c_2 - I(x))^2 (1 - u_1(x)) u_2(x) \, dx \\ & + \int_\Omega (c_3 - I(x))^2 (1 - u_1(x))(1 - u_2(x)) \, dx \Bigg\}. \end{aligned}$$

We see that it is straightforward to apply our general framework to the Vese-Chan functional (18). Indeed, the previous minimization problem can be written

$$\min_{\overrightarrow{u} \colon \Omega \to \{0,1\}^2} \left\{ \sum_{i=1}^{2} \int_\Omega |\nabla u_i| \, dx + \int_\Omega \rho(x, u_1(x), u_2(x)) \, dx \right\},$$

where

$$\begin{aligned} \rho(x, u_1(x), u_2(x)) = {} & (c_0 - I(x))^2 u_1(x) u_2(x) + (c_1 - I(x))^2 u_1(x)(1 - u_2(x)) \\ & + (c_2 - I(x))^2 (1 - u_1(x)) u_2(x) + (c_3 - I(x))^2 (1 - u_1(x))(1 - u_2(x)). \end{aligned}$$

In the original work [32], a solution was obtained using gradient descent on the time-dependent Euler-Lagrange equations of the energy functional,

yielding a local solution. On the other hand, using our convex technique, we may attempt to find a global solution and use straightforward calculations provided by Theorem 6 to determine whether our algorithm was successful. We should mention that the recent work of [3] proposes a graph-cut method in the discrete setting to globally solve the Vese-Chan model. We elect to develop our method in the spatially continuous setting.

Let us examine the extent to which our method will yield global solutions through some numerical examples. In Figure 4, we demonstrate our method on a synthetic image with noise taken from the original work of Vese and Chan [32, p. 17, Figure 5(b)]. The authors remark that the initial condition in Figure 4 causes their algorithm to compute only a local minimum, but that other initial conditions result in the global solution. On the other hand, our algorithm computes the global solution (or at least a very close approximation), no matter what initial condition is provided. The global minimizer for our algorithm ϕ^* gives the energy value 0.213989. which becomes 0.214023 after thresholding at 1. In contrast, The energy value for the obtained minimizer from the original Vese-Chan method is 0.223683.

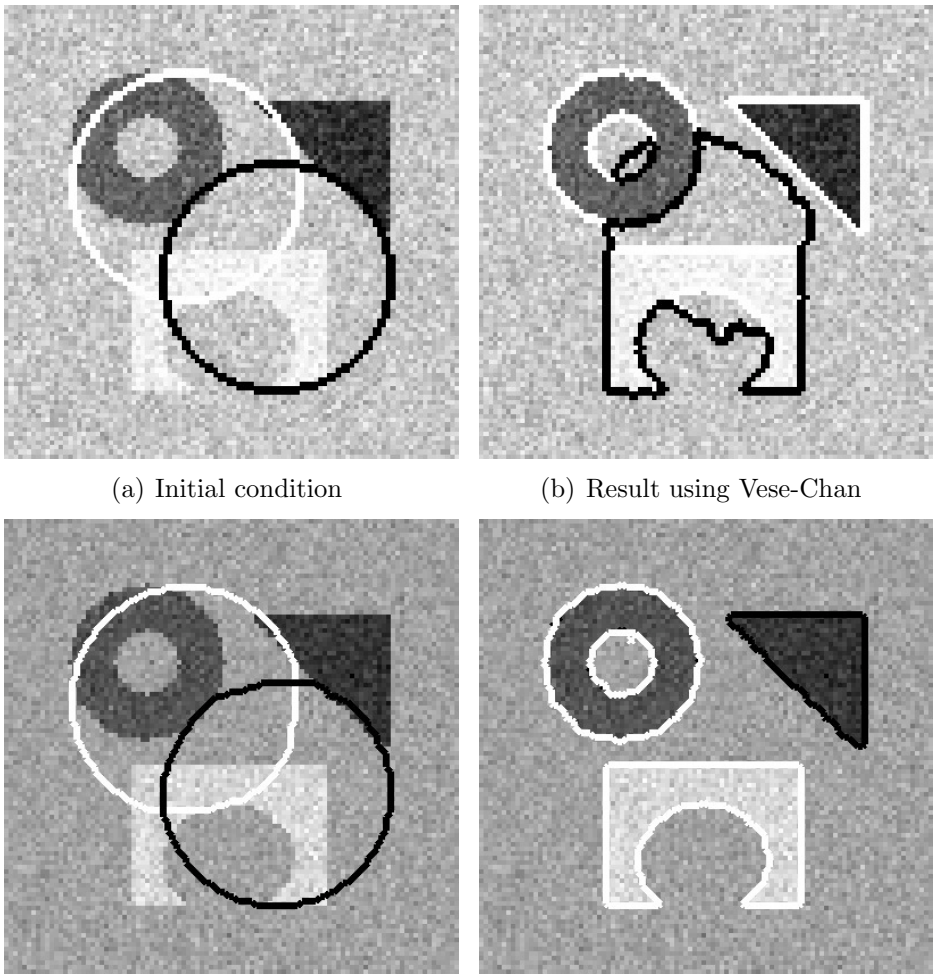
In Figure 5, we show more experimental results for this synthetic image. The histogram in Figure 5(a) shows that the obtained minimizer ϕ^* is close to binary. Furthermore, we see from Figure 5(b) that nearly all of the pixels x in the image are such that $\phi^*(x, \cdot)$ is a box function. The ratio of pixels which are non-box functions to the total number of pixels in the image is 0.0012.

We also demonstrate the effectiveness of our algorithm on a medical image segmentation application in Figure 6. Like the case of the synthetic segmentation example, our method is able to accurately segment the brain image into four phases. Again we see that the histogram of the minimizer ϕ^* is close to binary and the number of pixels where ϕ^* is not a box function is relatively small. The ratio of non-box function points to total points in ϕ^* in this case is 0.0141.

The Vese-Chan representation corresponds exactly to piecewise constant Mumford-Shah if and only if the zero-level sets do not overlap nontrivially, i.e.,

$$\mathcal{H}^{d-1}\left(\bigcap_{k=1}^{m}\{x \in \Omega\colon \varphi_k(x) = 0\}\right) = 0,$$

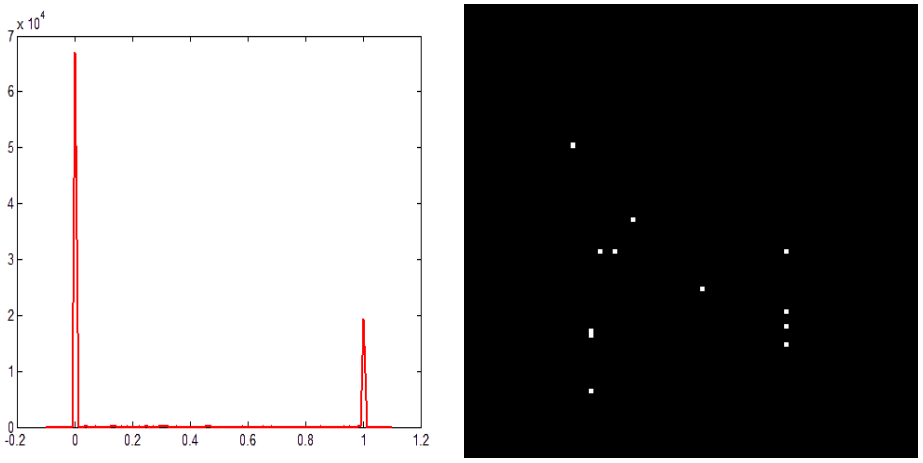
where $\mathcal{H}^{d-1}$ denotes $(d-1)$-dimensional Hausdorff measure. In particular, consider the triple junction example shown in Figure 7(a), the representa-



(a) Initial condition (b) Result using Vese-Chan

(c) Initial condition (d) Result using our method

Figure 4: A comparison of our algorithm and the original Vese-Chan method. For some images and initial conditions, the Vese-Chan method computes only a local minimum, whereas our algorithm is guaranteed to find the global solution under the conditions of Theorem 6.



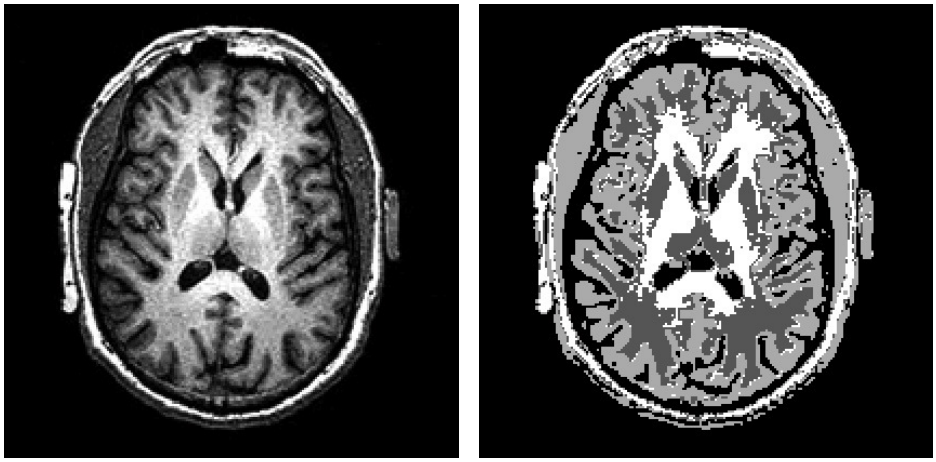
(a) Histogram (b) Indicator of non-box functions

Figure 5: (a) Histogram of minimizer ϕ^* of the functional F for the synthetic image in Figure 4. (b) A display indicating the spatial points $x \in \Omega$ for which $\phi^*(x, \cdot)$ is a box function; black (resp. white) indicate box (resp. non-box) funtions.

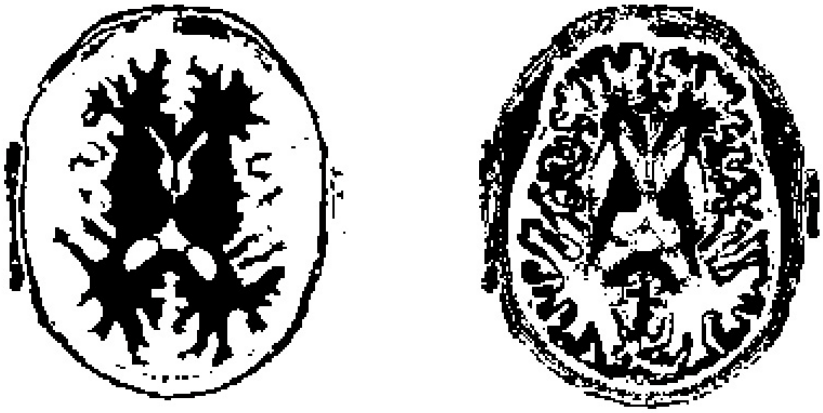
tion uses two functions whose zero-level sets necessarily have an intersection that coincides with exactly one of the boundaries. The effect is a non-uniform weighting of the boundaries, causing an incorrect segmentation, and thus the model cannot be used to segment the triple junction. We show the results in Figure 7. Notice in (c) and (d) that the boundaries determined by the functions u_1 and u_2 overlap non-trivially on the horizontal portion on the right side of the segmentation result. This means that portion of the boundary is weighted by a factor of two, causing the functional to be minimized when that boundary is shorter relative to the other boundaries. On the other hand, the Vese-Chan framework can correctly represent a quadruple junction; the results of our method are shown in Figure 8.

Let us examine the extent to which the hypotheses of Theorem 6 are satisfied. More precisely, for each of the numerical examples, we check whether the certificate condition $F(\phi^*) = F(1_{\{\phi^* \geq 1\}})$ holds. The results are shown in Table 1.

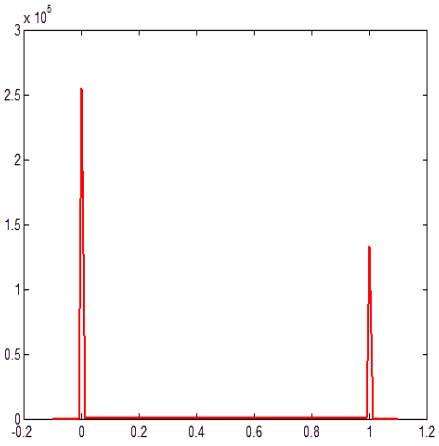
Finally, we show that under certain circumstances, we can prove that the vector-valued minimization method proposed in Section 2 is guaranteed to yield global solutions.

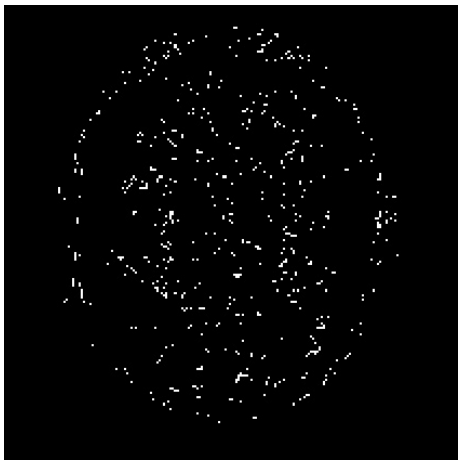


(a) Input (b) Result



(c) u_1 (d) u_2



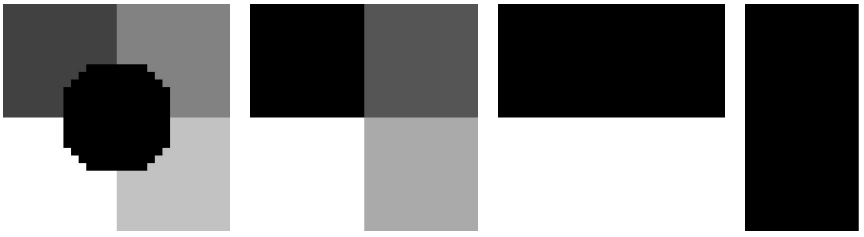


(e) Histogram (f) Box function indicator

Figure 6: (a)-(d): Segmentation of an MRI brain image into four phases using our method applied to the Vese-Chan model. (e) Histogram of minimizer ϕ^* (f) A display indicating the spatial points for which $1_{\{\phi^*(x,\cdot)\}}$ is a box function; black (resp. white) indicate box (resp. non-box) funtions.

(a) Input (b) Result (c) u_1 (d) u_2

Figure 7: Segmentation of a triple junctions using our method applied to the Vese-Chan model. A triple junction cannot be represented in the Vese-Chan framework so that the boundaries are evenly weighted, so the desired 120° intersection is not obtained.



(a) Input (b) Result (c) u_1 (d) u_2

Figure 8: Segmentation of a quadruple junction using our method applied to the Vese-Chan model. Unlike a triple junction, a quadruple junction can be represented so that the boundaries are evenly weighted, resulting in the correct segmentation.

Image	$F(\phi^*)$	$F(1_{\{\phi^* \geq 1\}})$
Synthetic	0.2139	0.2140
Brain	0.38	0.45
Quadruple	0.054	0.054
Triple	7.087	7.088

Table 1: Numerical values for the certificate condition of Theorem 6.

Proposition 7. *Suppose that $\rho \geq 0$ and that for all $x \in \Omega$, the minimization problem*

$$\min_{\overrightarrow{\gamma} \in \Gamma} \rho(x, \overrightarrow{\gamma})$$

has a unique solution $\overrightarrow{u}(x)$. Then

$$\min_{\phi \in X} F(\phi) = \int_\Omega \sum_{\overrightarrow{\gamma} \in \Gamma} (-1)^m \rho(x, \overrightarrow{\gamma}) D^m \phi$$

has a unique minimizer $\phi^* = 1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}}(x, \overrightarrow{\gamma})$.

Proof. The constraint $\phi \in X$ is pointwise in x as is the functional F, so it suffices to minimize

$$\sum_{\overrightarrow{\gamma} \in \Gamma} (-1)^m \rho(x, \overrightarrow{\gamma}) D^m \phi(x, \overrightarrow{\gamma})$$

for fixed x. Since $\sum_{\overrightarrow{\gamma}} (-1)^m D^m \phi(x, \overrightarrow{\gamma}) = 1$ and $(-1)^m D^m \phi \geq 0$, we see that $F = \int_\Omega F_x \, dx$ with $F_x = \sum_{\overrightarrow{\gamma}} c_{\overrightarrow{\gamma}} \rho(x, \overrightarrow{\gamma})$, where the coefficients $c_{\overrightarrow{\gamma}} := (-1)^m D^m \phi(x, \overrightarrow{\gamma})$ form a convex combination of $\rho(x, \overrightarrow{\gamma})$. Since, by assumption, the function $\rho(x, \overrightarrow{\gamma})$ has a unique minimizer $\overrightarrow{u}(x)$ for each $x \in \Omega$, it follows that the minimizer ϕ^* of F over all $\phi \in X$ has the property that $(-1)^m D^m \phi^*(x, \overrightarrow{u}(x)) = 1$ and $(-1)^m D^m \phi^*(x, \overrightarrow{\gamma}) = 0$ for all $\overrightarrow{\gamma} \neq \overrightarrow{u}(x)$, which implies that ϕ^* is the box function $1_{\{\overrightarrow{u} \succeq \overrightarrow{\gamma}\}}$. □

Of course, the total variation regularization is an important aspect of the segmentation model. In practice, there is often a parameter $\mu > 0$ in the Vese-Chan model that multiplies the regularization term, which we omitted from our description of the minimization problem for ease of presentation. However, it is worthwhile to investigate the effect changing this parameter has on our algorithm. For the brain image experiment, we used our algorithm for different values of μ. The results are shown in Table 2.

We see that as the magnitude of the total variation term increases, the amount by which the certificate condition fails increases. For images without much noise, accurate segmentations may be obtained with relatively small values of μ. As μ gets small, we can be more confident our algorithm produces a global minimizer of the problem.

μ	$F(\phi^*)$	$F(1_{\{\phi^* \geq 1\}})$	$F(1_{\{\phi^* \geq 1\}}) - F(\phi^*)$
0	0.432	0.432	0
0.5	0.384	0.453	0.069
1	0.374	0.478	0.104
2	0.399	0.537	0.138

Table 2: Numerical values for the certificate condition of Theorem 6 as the regularization parameter μ changes.

4.2 Piecewise Constant Mumford-Shah

We now use the general framework in Section 2 to develop a method to globally solve the piecewise constant Mumford-Shah problem (17). We use the framework of Lie et al. [18] to represent (17) as a minimization problem over a function $u \colon \Omega \to \{0, \ldots, n-1\}$ with the property that $u = i$ on Ω_i. Then (17) becomes

$$\min_{u \colon \Omega \to \{0,\ldots,n-1\}} \left\{ \sum_{i=0}^{n-1} \int_\Omega |\nabla \psi_i(u)| + \psi_i(u) g_i(x) \, dx \right\}, \tag{19}$$

where $g_i(x) = (c_i - I(x))^2$ and

$$\psi_i(u) := 1_{\{u=i\}} = \prod_{i \neq j} \frac{(u-j)}{(i-j)}.$$

We cannot simply apply our framework to (19) in the case $m = 1$ (i.e., the scalar-valued case of Pock et al. [24]), due to the regularization terms involving $\int_\Omega |\nabla \psi_i(u)| \, dx$; see [6] for more discussion.

Instead, we use a splitting technique. For $i = 0, \ldots, n-1$, we let $v_i = \psi_i(u)$ to obtain the problem

$$\min_{u \colon \Omega \to \{0,\ldots,n-1\},\, v_i \colon \Omega \to \{0,1\}} \left\{ \sum_{i=0}^{n-1} \int_\Omega |\nabla v_i| + v_i g_i(x) \, dx \right\} \text{ s.t. } v_i = \psi_i(u). \tag{20}$$

The constraints $v_i = \psi_i(u)$ may be enforced using an augmented Lagrangian approach [4]. For each $i = 0, \ldots, n-1$, choose a sequence of multipliers $\{\lambda_i^{(j)} \in L^2(\Omega)\}_{j=1}^\infty$ and penalty parameters $\{r_i^{(j)} > 0\}_{j=1}^\infty$ such that $\{\lambda_i^{(j)}\}_{j=1}^\infty$

is bounded and $r_i^{(j)} \to \infty$ as $j \to \infty$. For each j, let $(u^{(j)}, v_0^{(j)}, \ldots, v_{n-1}(j))$ be a global solution of

$$\min_{u,v_0,\ldots,v_{n-1}} \sum_{i=0}^{n-1} \int_\Omega |\nabla v_i| + v_i g_i(x) + \lambda_i^{(j)}(v_i - \psi_i(u)) + r_i^{(j)}(v_i - \psi_i(u))^2 \, dx, \quad (21)$$

where the minimization is taken over $u\colon \Omega \to \{0, \ldots, n-1\}$ and $v_i\colon \Omega \to \{0, 1\}$. Then $u^{(j)}$ converges to the solution of (17); see [6] for a proof (based on [4]) in a similar setting. Hence, it suffices to globally solve (21) for fixed j.

Observe that (21) can be written as

$$\min_{u\colon \Omega \to \{0,\ldots,n-1\}, v_i\colon \Omega \to \{0,1\}} \sum_{i=0}^{n-1} \int_\Omega |\nabla v_i| \, dx + \int_\Omega \rho(x, u, v_0, \ldots, v_{n-1}) \, dx$$

with

$$\rho(x, u, v_0, \ldots, v_{n-1}) = \sum_{i=0}^{n-1} v_i g_i(x) + \lambda_i^{(j)}(v_i - \psi_i(u)) + r_i^{(j)}(v_i - \psi_i(u))^2,$$

to which may apply the results established in Section 2. When the hypotheses of Theorem 6 are met, we can guarantee a global solution of (21), which yields a sequence converging to a global solution of (17).

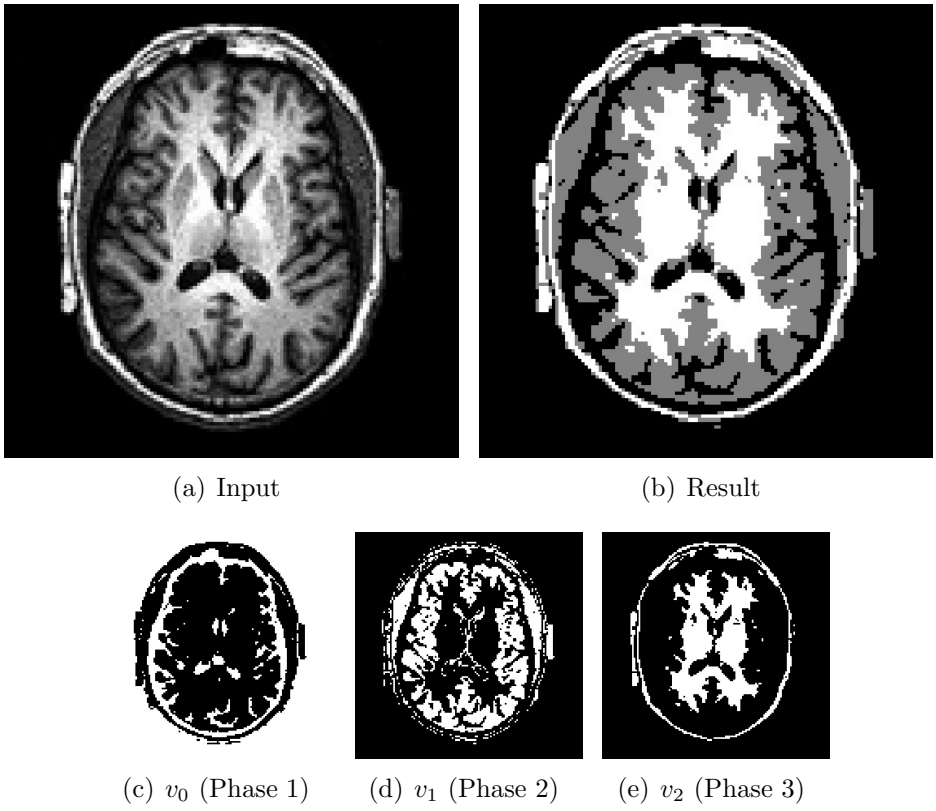
We now consider some numerical examples. In Figure 9, we show our method on the brain MRI image we considered in the previous subsection, except here we segment the image into three phases.

Unlike the case when our algorithm applied to the four-phase Vese-Chan model for this image, we find that the certificate condition holds exactly. That is, for the energy

$$\begin{aligned} F(\phi) = &\sum_{\ell=1}^{n-1} \int_\Omega |\nabla \phi(x, \ell\overrightarrow{e_1})| \, dx + \sum_{i=2}^{4} \int_\Omega |\nabla \phi(x, \overrightarrow{e_i})| \, dx \\ &+ \int_\Omega \sum_{\overrightarrow{\gamma} \in \Gamma} \rho(x, \overrightarrow{\gamma}) D^4 \phi(x, \overrightarrow{\gamma}) \, dx, \end{aligned}$$

where

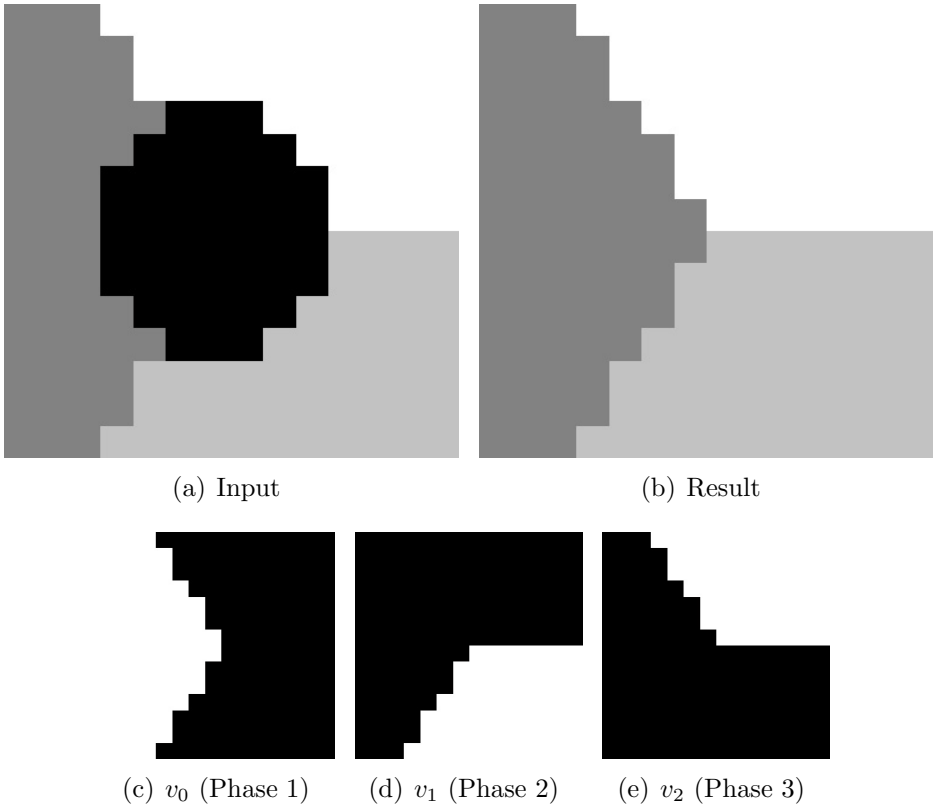
$$\rho(x, u_1, u_2, u_3, u_4) = \sum_{i=0}^{2} u_{i+2}(c_i - I(x))^2 + \lambda_i(u_{i+2} - \psi_i(u_1)) + r_i(u_{i+2} - \psi_i(u_1))^2,$$



(a) Input (b) Result

(c) v_0 (Phase 1) (d) v_1 (Phase 2) (e) v_2 (Phase 3)

Figure 9: Segmentation of an MRI brain image into three phases using the method of Section 4.2.



(a) Input (b) Result

(c) v_0 (Phase 1) (d) v_1 (Phase 2) (e) v_2 (Phase 3)

Figure 10: The method of Section 4.2 applied to the triple junction example, obtained by thresholding at $\frac{1}{2}$.

the minimizer ϕ^* of F over X satisfies the condition $F(\phi^*) = F(1_{\{\phi^* \geq 1\}})$, which both are equal to 0.33. In fact, in this case, the function ϕ^* itself is a box function, whose principal vertex corresponds to the minimizer of the original problem.

We also demonstrate the method developed in this section on the synthetic triple junction example we considered in Section 4.1. The results are shown in Figure 10 and should be compared with the results in Figure 7. We know that the piecewise constant Mumford-Shah model has the desired triple junction as its minimizer, while the model of Vese-Chan does not. The results of our method on the triple junction example are interesting. In this case, the certificate condition fails miserably: the energy of $1_{\{\phi^* \geq 1\}}$ is an astronomical 148 compared with the 0.32 of the optimal solution. On the other

(a) Result for threshold $t = 0.1$. (b) Result for threshold $t = 1$.

Figure 11: The method of Section 4.2 yields suboptimal results on the triple junction example for certain thresholds.

hand, in this case $1_{\{\phi^* \geq \frac{1}{2}\}}$ is a box function that equals the optimal solution. For comparison, we show some results for other thresholds in Figure 11 that yield suboptimal results.

We cannot fully explain why thresholding at $\frac{1}{2}$ provides the global solution for this example. But it is interesting to examine the values of the minimizer at spatial points near the triple junction intersection. For example, we took a point x near this intersection and found the following values of the minimizer: $\phi^*(x, 0, 1, 0, 0) = 0.5966, \phi^*(x, 0, 0, 1, 0) = 0.0561, \phi^*(x, 0, 0, 0, 1) = 0.3439$. The sum of these values is approximately 1, showing that a convex simplex constraint similar to that of the other relaxation methods [6, 33, 17] is appearing implicitly in our model. We leave further examination of this phenomenon for future work.

5 Conclusion

In this paper, we have established a general framework for solving certain vector-valued non-convex minimization problems that arise in image processing and computer vision applications. The method reformulates the non-convex functional as an equivalent convex functional in a higher dimensional space, where then a specific convex relaxation taken. Under certain condi-

tions, a global minimizer of the original problem can be guaranteed. More precisely, we propose to solve non-convex minimization problems of the form

$$\min_{\overrightarrow{u}} E(\overrightarrow{u}) := \sum_{i=1}^{m} \int_{\Omega} |\nabla u_i| \, dx + \int_{\Omega} \rho(x, \overrightarrow{u}(x)) \, dx \tag{22}$$

by reformulating the problem as

$$\min_{\phi = 1_{\{\overrightarrow{u} \geq \overrightarrow{\gamma}\}}} \left\{ \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \int_{\Omega} |\nabla \phi(x, \ell \overrightarrow{e_i})| \, dx + (-1)^m \sum_{\overrightarrow{\gamma} \in \Gamma} \int_{\Omega} \rho(x, \overrightarrow{\gamma}) D^m_{1,\ldots,m} \phi \, dx \right\}.$$

We then relax the problem in a specific way to obtain a convex minimization problem. The relaxation is shown to be exact under certain circumstances, which are explained in detail. The most useful of these conditions is the equality $F(\phi^*) = F(1_{\{\phi^* \geq 1\}})$, where F is the functional of the reformulated problem above and ϕ^* is its minimizer over the relaxed set. In this case, our method can guarantee a global minimizer.

However, the luxury of an optimal solution comes at a considerable price. Indeed, the difficulty of the non-convexity has been transformed into a difficulty of dimensionality. Even for some modest non-convex problems, the convex counterpart can be so computationally burdensome that the method at present may prove to be impractical. Still, with our framework in hand, further research on the topic may uncover improvements that allow our method to be as meaningful in practice as it is in theory.

In Section 1, we made the remark that the method of Pock et al. [24], which can be seen as the $m = 1$ case in our framework, is also related to calibration theory [1] and a general framework [7] whereby functionals of the form

$$\int_{\Omega} f(x, u(x), \nabla u) \, dx$$

with f convex in its last argument, can be reformulated in a higher dimension as the convex functional

$$\sup_{\overrightarrow{p} \in K} \int_{\Omega \times \mathbb{R}} \overrightarrow{p} \cdot \nabla 1_{u > \gamma},$$

with K a convex set that can be written in terms of the convex conjugate f^* (taken with respect to its last argument). The theory has been generalized to the case in which u is vector-valued rather than scalar valued [20]. It would be interesting to examine the relationship this theory has with our framework. We leave this for future work.

Acknowledgments

The authors would like to thank Tom Goldstein for very helpful discussion.

References

[1] G. Alberti, G. Bouchitt, and G. D. Maso. The calibration method for the mumford-shah functional. *Comptes Rendus de l'Acadmie des Sciences - Series I - Mathematics*, 329(3):249 – 254, 1999.

[2] K. J. Arrow, L. Hurwicz, and H. Uzawa. *Studies in linear and non-linear programming*. With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow. Stanford Mathematical Studies in the Social Sciences, vol. II. Stanford University Press, Stanford, Calif., 1958.

[3] E. Bae and X.-C. Tai. Efficient global optimization for the multiphase chan-vese model of image segmentation by graph cuts. In *EMMCVPR*, 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] E. S. Brown, T. F. Chan, and X. Bresson. Convex formulations for piecewise constant mumford-shah image segmentation. *UCLA CAM Report 09-66*, 2009.

[7] A. Chambolle. Convex representation for lower semicontinuous envelopes of functionals in L^1. *J. Convex Anal.*, 8(1):149–170, 2001.

[8] 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.

[9] 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.

[10] 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.

[11] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for tv minimization. *UCLA CAM Report 09-67*, 2009.

[12] W. Fleming and R. Rishel. An integral formula for total gradient variation. *Archiv der Mathematik*.

[13] M. Giaquinta, G. Modica, and J. Sou?ek. Cartesian currents in the calculus of variations. ii, volume 37 of ergebnisse der mathematik und ihrer grenzgebiete. 3. folge. a. In *Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics*. Springer-Verlag, 1998.

[14] T. Goldstein, X. Bresson, and S. Osher. Global minimization of markov random fields with applications to optical flow. *UCLA CAM Report 09-77*, September, 2009.

[15] B. K. P. Horn and B. G. Schunck. Determining optical flow. *ARTIFICAL INTELLIGENCE*, 17:185–203, 1981.

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

[17] 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.

[18] 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.

[19] E. H. Lieb and M. Loss. *Analysis*. American Mathematical Society, Providence, RI, 2001.

[20] M. G. Mora. The calibration method for free-discontinuity problems on vector-valued maps. *J. Convex Anal.*, 9(1):1–29, 2002.

[21] 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.

[22] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12 – 49, 1988.

[23] T. Pock, A. Chambolle, H. Bischof, and D. Cremers. An algorithm for minimizing the mumford-shah functional. In *IEEE Conference on Computer Vision (ICCV)*, 2009.

[24] 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.

[25] L. D. Popov. A modification of the Arrow-Hurwitz method of search for saddle points. *Mat. Zametki*, 28(5):777–784, 803, 1980.

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

[27] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[28] R. T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2):183–238, 1993.

[29] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

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

[31] H. B. A. C. Thomas Pock, Daniel Cremers. Global solutions of variational models with convex regularization. *Technical Report, Institute for Computer Graphics and Vision, Graz University of Technology*, 2009.

[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.