Int J Comput Vis (2012) 98:103–121
DOI 10.1007/s11263-011-0499-y

Completely Convex Formulation of the Chan-Vese Image Segmentation Model

Ethan S. Brown · Tony F. Chan · Xavier Bresson

Received: 14 July 2010 / Accepted: 27 September 2011 / Published online: 20 October 2011

Abstract The active contours without edges model of Chan and Vese (IEEE Transactions on Image Processing 10(2):266–277, 2001) is a popular method for computing the segmentation of an image into two phases, based on the piecewise constant Mumford-Shah model. The minimization problem is non-convex even when the optimal region constants are known *a priori*. In (SIAM Journal of Applied Mathematics 66(5):1632–1648, 2006), Chan, Esedoḡlu, and Nikolova provided a method to compute global minimizers by showing that solutions could be obtained from a convex relaxation. In this paper, we propose a convex relaxation approach to solve the case in which both the segmentation and the optimal constants are unknown for two phases and multiple phases. In other words, we propose a convex relaxation of the popular K-means algorithm. Our approach is based on the vector-valued relaxation technique developed by Goldstein et al. (UCLA CAM Report 09-77, 2009) and Brown et al. (UCLA CAM Report 10-43, 2010). The idea is to consider the optimal constants as functions subject to a constraint on their gradient. Although the proposed relaxation technique is not guaranteed to find exact global minimizers of the original problem, our experiments show that our method computes tight approximations of the optimal solutions. Particularly, we provide numerical examples in which our method finds better solutions than the method proposed by Chan et al. (SIAM Journal of Applied Mathematics 66(5):1632–1648, 2006), whose quality of solutions depends on the choice of the initial condition.

Keywords Image segmentation · Chan-Vese model · Convex relaxation · Level set method · Vector-valued functional lifting · K-means

1 Introduction

Image segmentation is a fundamental problem in image processing and computer vision. The goal is to divide the image into regions that belong to distinct objects in the given image. Many successful approaches to image segmentation involve variational models and partial differential equations. These models are well suited to impose regularity constraints for the solutions.

The Mumford-Shah model (Mumford and Shah 1989) is one of the most influential image segmentation models. In particular, the two-phase, piecewise constant case has been extensively studied. Under this model, the image is partitioned into two phases (i.e., object and background) and is approximated by a function that takes only two values. The algorithm of Chan and Vese (2001) is one of the most widely used methods for two-phase image segmentation. Their method, also known as active contours without edges, is based on techniques of curve evolution and the level set method (Osher and Sethian 1988).

However, as is the case for many variational image processing models, the energy functional to be minimized is non-convex and therefore has local minima. This is a serious

E.S. Brown
Department of Mathematics, University of California
Los Angeles, Los Angeles, USA
e-mail: ethan@math.ucla.edu

T.F. Chan
Hong Kong University of Science and Technology, Hong Kong, Hong Kong
e-mail: tonyfchan@ust.hk

X. Bresson (✉)
Department of Computer Science, City University of Hong Kong, Hong Kong, Hong Kong
e-mail: xbresson@cityu.edu.hk

difficulty because the local minima of segmentation models often provide poor results. The Chan-Vese method, as well as many other solution techniques, is based on gradient descent and is prone to getting stuck in such local minima. Consequently, to obtain satisfactory results, the initialization in these algorithms is very important.

The important work of Chan et al. (2006) showed that this non-convex optimization problem can be solved by a convex relaxation method when the two piecewise constant values are known. Namely, the source of the non-convexity is the non-convex collection of characteristic functions of sets. Based on the observations of Strang (1983), the idea is to relax this constraint in such a way that the characteristic function minimizers can be obtained from minimizers of the relaxed problem by a simple thresholding procedure. This allows the non-convex Chan-Vese problem to be globally solved using standard convex minimization methods. Note also that the idea of using characteristic functions rather than level set functions to represent regions in geometric problems has been first proposed by Lie et al. (2006).

When the piecewise constant values are unknown, the situation becomes more difficult. It is no longer possible to apply the aforementioned method to globally solve the two-phase piecewise constant segmentation problem. Instead, the standard approach would be use to alternate between globally solving the problem for fixed constant values, and updating these values. Unfortunately, this removes any guarantee that a global solution is obtained. There is extensive literature on global methods for segmentation in the case in which the values are known (most notably, attempts to extend the ideas of Chan et al. (2006) to the multiphase case (Zach et al. 2008; Lellmann et al. 2009; Chambolle et al. 2008; Brown et al. 2009, 2010; Bae et al. 2009, 2010; Lellmann et al. 2009; Lellmann and Schnörr 2010; Bae and Tai 2009; Goldluecke and Cremers 2010). However, except the recent work of Strandmark et al. (2009), there is surprisingly little work done in the direction of global methods in which the values are unknown. This is the subject of our present work.

In this paper, we propose a completely convex formulation of the Chan and Vese model, i.e. we look for solutions of the Chan and Vese model when the constant values for each phase are not known *a priori*. Our approach relies on a convex relaxation method for solving total variation-based vector-valued minimization problems introduced by Goldstein et al. (2009) for image registration problems and expanded upon in Brown et al. (2010) for multi-phase image segmentation problems, based originally on the ideas of Pock et al. (2008) and Ishikawa (2003). The idea is to embed the problem in a higher-dimensional space and then perform a convex relaxation. To employ this vector-valued optimization technique, we consider the piecewise constant values in each phase as functions whose gradients vanish everywhere. From the perspective of a two-phase segmentation problem, the unknown vector is a triple consisting of one binary-valued function that describes the segmentation, and two functions (enforced to be constant) that describe the piecewise constant value in each phase. Once we develop our method, we demonstrate examples where our approach obtains solutions close to the global solution but the alternating method of Chan et al. (2006) does not (because it strongly depends on the choice of the initial condition). Although the proposed relaxation technique is not guaranteed to find exact global minimizers of the original Chan-Vese problem, our experiments show that our method computes tight approximations of the optimal solutions. Indeed, we propose a certificate that guarantees the relaxed solution to be the exact, i.e. the solution of the original non-convex problem. Our method provides solutions close to the optimal solutions in the sense that the error on the optimality certificate is only 0.01–1.8% for all experiments.

The outline of the paper is as follows. In Sect. 2, we introduce the segmentation problem and describe the convex relaxation method of Chan et al. (2006) that solves the problem when the optimal constants are known. Next, in Sect. 3, we review the main ideas of the total variation-based vector-valued convex relaxation method described in Brown et al. (2010), which serves as our primary tool to develop our convex relaxation method. The central idea of the paper is in Sect. 4, where we propose a method to globally solve the Chan-Vese problem even when the region constants are unknown. We give experimental results in Sect. 5 that, in particular, show that our method can obtain better minimizers than the previous method (Chan et al. 2006). Finally, we end the paper with a concluding section. Throughout the paper, the mathematical space that we consider is the space of functions with bounded variation (BV) (we refer the reader e.g. to Ambrosio et al. 2000 for more details).

2 Chan-Vese with Known Constants

The problem considered in this paper is the piecewise constant two-phase Mumford-Shah segmentation model (Mumford and Shah 1989). The problem can be expressed as the minimization problem

$$\min_{\Sigma \subset \Omega,\, c_1, c_2 \in \mathbb{R}} MS(\Sigma, c_1, c_2), \tag{1}$$

where

$$\begin{aligned} MS(\Sigma, c_1, c_2) := \mathrm{Per}(\Sigma; \Omega) &+ \lambda \int_{\Sigma} (c_1 - I(x))^2 \, dx \\ &+ \lambda \int_{\Omega \setminus \Sigma} (c_2 - I(x))^2 \, dx, \end{aligned} \tag{2}$$

where $\mathrm{Per}(\Sigma;\Omega)$ denotes the perimeter of the set $\Sigma \subset \Omega$ (throughout the paper, the set $\Omega \subset \mathbb{R}^n$ is supposed to be open and bounded with $\partial\Omega$ Lipschitz; see Evans and Gariepy 2000). In other words, we seek the best approximation to the given image $I: \Omega \to \mathbb{R}$ among all functions that can take only two values. The unknowns of the optimization problem are the values, c_1 and c_2, and the sets where each value is taken, Σ and $\Omega \setminus \Sigma$. As usual, there is a regularization term that penalizes size of the boundary $\partial\Sigma$ that separates the two phases. In the seminal work (Chan and Vese 2001), Chan and Vese proposed a level-set based algorithm for solving (1). Due to the popularity of their algorithm, the problem (1) itself is often referred to Chan-Vese and we too often follow this precedent. In their approach, the boundary $\partial\Sigma$ is represented by the zero-level set of a function $\varphi: \Omega \to \mathbb{R}$. In terms of φ, the energy (2) can be written

$$CV(\varphi, c_1, c_2) = \int_\Omega |\nabla H(\varphi)| + \lambda \int_\Omega H(\varphi)(c_1 - I(x))^2\,dx + \lambda \int_\Omega (1 - H(\varphi))(c_2 - I(x))^2\,dx, \tag{3}$$

where the gradient operator ∇ is taken in the distributional sense. To compute the variation of (3) with respect to φ, a smooth regularization H_ϵ of the Heaviside function H is introduced to obtain the approximation

$$\begin{aligned} CV_\epsilon(\varphi, c_1, c_2) = &\int_\Omega |\nabla H_\epsilon(\varphi)| \\ &+ \lambda \int_\Omega H_\epsilon(\varphi)(c_1 - I(x))^2\,dx \\ &+ \lambda \int_\Omega (1 - H_\epsilon(\varphi))(c_2 - I(x))^2\,dx. \end{aligned} \tag{4}$$

Variations of (4) lead to the gradient descent scheme

$$\varphi_t = H'_\epsilon(\varphi)\left\{\mathrm{div}\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right) - \lambda((c_1 - I(x))^2 - (c_2 - I(x))^2)\right\}. \tag{5}$$

We return now to the minimization problem (1). Observe that if Σ is fixed, the values of c_1 and c_2 that minimize $MS(\Sigma,\cdot,\cdot)$ are simply

$$\begin{aligned} c_1 &= \frac{1}{|\Sigma|}\int_\Sigma I(x)\,dx, \\ c_2 &= \frac{1}{|\Omega \setminus \Sigma|}\int_{\Omega\setminus\Sigma} I(x)\,dx, \end{aligned} \tag{6}$$

that is, c_1 and c_2 are the mean values of the image $I(x)$ in the regions Σ and $\Omega \setminus \Sigma$, respectively. A natural method to minimize MS is an alternating scheme that first computes the values c_1 and c_2 using (6), and second solves the minimization problem $MS(\cdot, c_1, c_2)$ for the unknown set Σ. But even the minimization $MS(\cdot, c_1, c_2)$ for fixed c_1, c_2 is difficult because it is a non-convex problem.

Chan et al. (2006) proposed a convex relaxation method to solve this non-convex problem. The authors observed that (5) has the same stationary solutions as

$$\varphi_t = \mathrm{div}\left(\frac{\nabla\varphi}{|\nabla\varphi|}\right) - \lambda((c_1 - I(x))^2 - (c_2 - I(x))^2), \tag{7}$$

which in turn is the gradient descent for the energy

$$\int_\Omega |\nabla\varphi| + \lambda\int_\Omega ((c_1 - I(x))^2 - (c_2 - I(x))^2)\varphi\,dx. \tag{8}$$

Since this energy is homogeneous of degree 1, it does not have a minimizer in general, but does when the minimization of φ is restricted to the interval $[0, 1]$. Following the ideas of Strang (1983), they were able to show the following result.

Theorem 1 [Chan et al. 2006, Theorem 2] *Given fixed $c_1, c_2 \in \mathbb{R}$, a global minimizer for $MS(\cdot, c_1, c_2)$ can be found by conducting the convex minimization*

$$\min_{0\le u\le 1}\left\{E_{\mathrm{CEN}}(u) = \int_\Omega |\nabla u| + \lambda\int_\Omega u(x)(c_1 - I(x))^2 + (1 - u(x))(c_2 - I(x))^2\,dx\right\} \tag{9}$$

and setting $\Sigma = \{x \in \Omega: u(x) \ge t\}$ for almost any $t \in (0, 1]$.

Once we have a convex minimization problem, there is a wide catalogue of methods we can use to find a global minimizer. For example, (9) is equivalent to the unconstrained problem

$$\min_u \int_\Omega |\nabla u| + \int_\Omega \alpha\nu(u) + \lambda s(x)u\,dx, \tag{10}$$

where $\nu(\xi) = \max(0, 2|\xi - \frac{1}{2}| - 1)$, $\alpha > \frac{\lambda}{2}\|s\|_{L^\infty(\Omega)}$ and $s(x) = (c_1 - I(x))^2 - (c_2 - I(x))^2$. The first variation of (10) with respect to u leads to the Euler-Lagrange equation

$$\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right) - \lambda s(x) - \alpha\nu'(u) = 0, \tag{11}$$

which can be solved using a straightforward gradient descent scheme, using a regularized version of ν to smooth its non-differentiable points at 0 and 1, and using a regularized version of total variation to avoid degeneracy when $|\nabla u| = 0$. In fact, this method is both slow (being an explicit scheme) and inaccurate (due to the regularizations). More accurate and efficient algorithms for problems such as (10) were provided in Bresson et al. (2007), Goldstein et al. (2009), Yuan et al. (2010), El-Zehiry et al. (2007) and Chambolle and Pock (2010).

In summary, we have an alternating algorithm for solving the Chan-Vese problem (1), shown in Algorithm 1. Each step in the body of the while loop of the algorithm finds a global minimizer with the other variables fixed. Nevertheless, this alternating algorithm does not necessarily find a global minimizer of (1). To this end, we develop a different approach in the sections that follow.

Algorithm 1 Alternating Chan-Vese algorithm

$c_1 \leftarrow init_1$
$c_2 \leftarrow init_2$
while not converged **do**
  $\Sigma \leftarrow \operatorname{argmin}_\Sigma MS(\cdot, c_1, c_2)$
  $c_1 \leftarrow \frac{1}{|\Sigma|}\int_\Sigma I(x)\,dx$
  $c_2 \leftarrow \frac{1}{|\Omega\setminus\Sigma|}\int_{\Omega\setminus\Sigma} I(x)\,dx$
end while

3 Total Variation-Based Vector-Valued Minimization

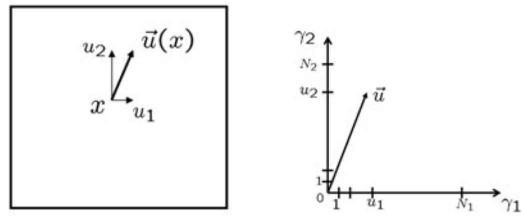
In this section, we describe a method to formulate a class of total variation-based non-convex optimization problems as equivalent convex optimization problems by embedding the problem into a higher dimensional space. The method was introduced in Goldstein et al. (2009), which generalized the ideas of Ishikawa (2003), Pock et al. (2008) to the case in which the unknown is vector-valued rather than scalar-valued. The approach was expanded upon in Brown et al. (2010), where it was applied to multi-phase image segmentation problems. In this paper, we will eventually use this minimization technique to formulate a convex relaxation of the Chan-Vese problem (1), and provide tight approximations of the optimal solutions.

3.1 Theory

The class of optimization problems we consider in this section are of the form

$$\min_{\vec{u}} E(\vec{u}) := \sum_{i=1}^{m}\int_\Omega |\nabla u_i| + \int_\Omega \rho(x, \vec{u}(x))\,dx, \tag{12}$$

where $\vec{u} = (u_1, \dots, u_m)\colon \Omega \to \Gamma := \{0, 1, \dots, N_1\} \times \dots \times \{0, 1, \dots, N_m\}$ (see Fig. 1). In other words, the unknown is a function $\vec{u} = (u_1, \dots, u_m)$ defined on a continuous domain $\Omega \subset \mathbb{R}^d$, and each of its components u_i takes values in the discrete set $\{0, 1, \dots, N_i\}$. 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 continuous and 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 convexity assumption on ρ; thus ρ may be non-convex.



(a) Domain Ω (b) Domain Γ

Fig. 1 (**a**) Image domain Ω. (**b**) Domain Γ with $m = 2$: at each pixel x in the image domain, the vector field $\vec{u} = (u_1, u_2)$ is discretized in the Γ domain

To embed (12) into a higher dimensional space, we introduce the function

$$1_{\{\vec{u}(x)\succeq\vec{\gamma}\}} := 1_{\{u_1\ge\gamma_1,\dots,u_m\ge\gamma_m\}}(x, \vec{\gamma}) = \begin{cases} 1 & \text{if } u_1 \ge \gamma_1, \dots, u_m \ge \gamma_m, \\ 0 & \text{otherwise.}\end{cases} \tag{13}$$

We call such a function a *box function*, since for fixed $x \in \Omega$, the set of points $\vec{\gamma}$ in the non-negative orthant of $\mathbb{R}^m$ where $1_{\{\vec{u}\succeq\vec{\gamma}\}}$ is equal to 1 is a hypercube. We call the point $\vec{u}$ 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 $\vec{u}$ and its associated box function. In particular, we may use the formula

$$u_i(x) = \sum_{\ell=1}^{N_i} 1_{\{\vec{u}(x)\succeq\vec{\gamma}\}}(x, \ell\vec{e_i}) \tag{14}$$

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

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

$$\tilde{\Gamma} := \{0, \dots, N_1 + 1\} \times \dots \times \{0, \dots, N_m + 1\}, \tag{15}$$

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

$$(D_i\phi)(x, \vec{\gamma}) = \begin{cases} 0 & \text{if } \gamma_i = N_i + 1, \\ \phi(x, \vec{\gamma} + \vec{e_i}) - \phi(x, \vec{\gamma}) & \text{otherwise.}\end{cases} \tag{16}$$

Finally, we will often work with the set

$$C = \{\phi\colon \Omega \times \widetilde{\Gamma} \to [0,1]\colon \phi(x, \vec{0}) = 1 \text{ and } \phi(x, \vec{\gamma}) = 0 \text{ whenever } \gamma_i = N_i + 1 \text{ for some } i\}. \tag{17}$$

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

We point out a few properties of box functions in order to rewrite (12) in terms of the box function (13). First, observe that

$$1_{\{u_i \geq \gamma_i\}} = 1_{\{\vec{u} \geq \vec{\gamma}\}}(x, \gamma_i \vec{e_i}) \tag{18}$$

for each $i = 1, \dots, m$. Thus, by the coarea formula (Federer 1959; Fleming and Rishel 1960),

$$\int_\Omega |\nabla u_i| = \sum_{\ell=1}^{N_i} \int_\Omega \left|\nabla 1_{\{\vec{u} \geq \ell \vec{e_i}\}}\right|. \tag{19}$$

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

$$(D^m_{1,\dots,m} 1_{\{\vec{u} \geq \vec{\gamma}\}})(x, \vec{\gamma}) = \begin{cases} (-1)^m & \text{if } \vec{\gamma} = \vec{u}, \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

We illustrate this difference operator in Fig. 2.

Consequently,

$$(-1)^m \sum_{\vec{\gamma} \in \Gamma} \rho(x, \vec{\gamma}) D^m_{1,\dots,m} 1_{\{\vec{u} \geq \vec{\gamma}\}} = \rho(x, \vec{u}(x)), \tag{21}$$

and so we can rewrite problem (12), using the identities (19) and (21), as

$$\min_{\phi = 1_{\{\vec{u} \geq \vec{\gamma}\}}} F(\phi), \tag{22}$$

where

$$F(\phi) := \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \int_\Omega |\nabla \phi(x, \ell \vec{e_i})| + (-1)^m \sum_{\vec{\gamma} \in \Gamma} \int_\Omega \rho(x, \vec{\gamma}) D^m_{1,\dots,m} \phi \, dx. \tag{23}$$

That is, (12) 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 $\vec{u}$ (due to the function ρ), the reformulated objective function is convex in $\phi = 1_{\{\vec{u} \geq \vec{\gamma}\}}$. 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.

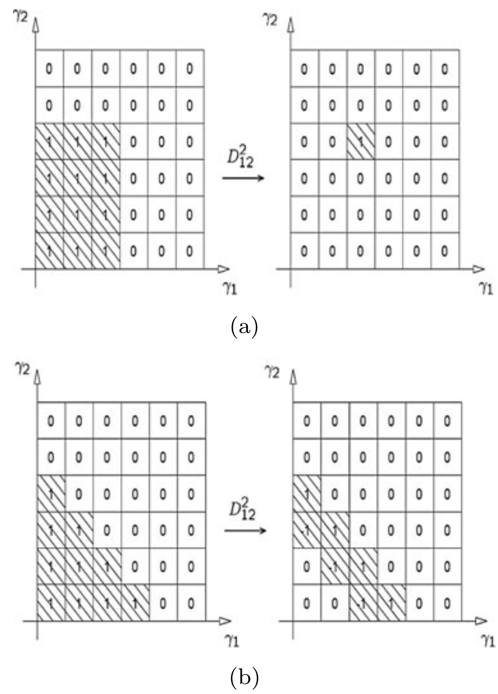


Fig. 2 An illustration of the difference operator $D^m_{1,\dots,m}$ when $m = 2$ applied to binary functions. In (**a**), we see that if ϕ is a box function, then the support of $D^m_{1,\dots,m}\phi$ is a single point (the principal vertex). On the other hand, in (**b**) we see that the support of $D^m_{1,\dots,m}\phi$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

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. We will develop a series of several preliminary results. The first result is that the functional F defined in (23) satisfies a generalized coarea formula of the form

$$F(\phi) = \int_0^1 F(1_{\{\phi \geq t\}})\, dt. \tag{24}$$

Lemma 2 *Let F be defined by (23). Then F satisfies the generalized coarea formula (24).*

Proof See the Appendix. □

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 3 *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 almost 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 See the Appendix. □

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 (22), and hence to (12). 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

$$X = \{\phi \in C\colon (-1)^m D^m_{1,\ldots,m} \phi \geq 0\}. \tag{25}$$

Observe that the set of all box functions is a subset of X. This leads to the following proposition.

Lemma 4 *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 (22) for all $t \in (0,1]$.*

Proof See the Appendix. □

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 (22) without appealing to the coarea formula (19).

Lemma 5 *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 (22).*

Proof See the Appendix. □

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 6 *For all $x \in \Omega$, suppose that $\phi\colon \Omega \times \widetilde{\Gamma} \to [0,1]$ lies in the set X defined in (25), i.e. ϕ satisfies the boundary conditions*

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

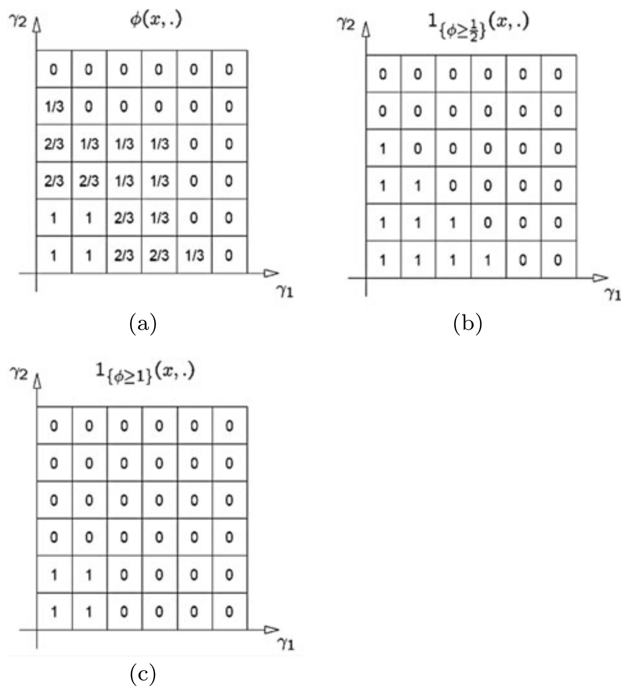


Fig. 3 The function $\phi(x,\cdot)$ in (**a**) satisfies the hypotheses of Lemma 6. However, in (**b**) we see that $1_{\{\phi \geq \frac{1}{2}\}}$ is not a box function. As shown in (**c**), $1_{\{\phi \geq 1\}}$ is a box function, as guaranteed by the lemma

and ϕ satisfies the difference condition

$$(-1)^m D^m_{1,\ldots,m} \phi \geq 0. \tag{26}$$

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

Proof See the Appendix. □

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 Fig. 3. We see that ϕ satisfies the hypotheses of Lemma 6, but that, for example, $1_{\{\phi \geq \frac{1}{2}\}}$ is not a box function. However, as guaranteed by the lemma, $1_{\{\phi \geq 1\}}$ 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. Goldstein et al. 2009, 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. Chambolle et al. 2008, 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 5 and Lemma 6, we can now state the main result of this section.

Theorem 7 *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*

$$\phi(x, \overrightarrow{0}) = 1 \quad \textit{and} \quad \phi(x, \gamma_1, \dots, \gamma_m) = 0 \quad \textit{whenever} \quad \gamma_i = N_i + 1 \textit{ for some } i \tag{27}$$

and

$$(-1)^m D^m_{1,\dots,m}\phi \geq 0. \tag{28}$$

Then

$$u_i(x) = \sum_{\ell=1}^{N_i} 1_{\{\phi^* \geq 1\}}(x, \ell\overrightarrow{e_i}) \tag{29}$$

is a minimizer of (12) *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 (12).

When $m = 1$, the results of this section reduce to the method proposed in Pock et al. (2008). 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 4 since we know the hypotheses are satisfied. However, when $m > 1$, this is no longer the case.

3.2 Algorithm

To conclude this section, we mention an algorithm to solve the minimization problem using the main theorem.

$$\min_{\phi\in X} F(\phi), \tag{30}$$

where F is given by (23). We write (30) as a saddle point problem

$$\min_{\phi\in C} \max_{|\overrightarrow{p_i}|\leq 1, (-1)^m p_\gamma \leq \rho} \left\{ \sum_{i=1}^m \sum_{\ell=1}^{N_i} \int_\Omega \overrightarrow{p_i} \cdot \nabla\phi(x, \ell\overrightarrow{e_i}) + (-1)^m \sum_{\overrightarrow{\gamma}\in\Gamma} p_\gamma D^m \phi \right\} \tag{31}$$

and solve it using a primal-dual algorithm. Recall the set C was defined in (17). 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.

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 (31) is thus

$$\min_{\phi^h\in C^h} \max_{\overrightarrow{p}\in D^h} \left\{ \sum_{i=1}^m \sum_{\ell=1}^{N_i} \sum_{x^h\in\Omega^h} \overrightarrow{p_i} \cdot \nabla^h \phi^h(x^h, \ell\overrightarrow{e_i}) + (-1)^m \sum_{\overrightarrow{\gamma}\in\Gamma} \sum_{x^h\in\Omega^h} p_\gamma D^m_{1,\dots,m}\phi^h \right\}, \tag{32}$$

where ∇^h is the discrete gradient operator,

$$C^h = \left\{\phi^h\colon \Omega^h \times \widetilde{\Gamma} \to [0,1]\colon \phi^h(x^h, \overrightarrow{0}) = 1 \text{ and } \phi^h(x^h, \overrightarrow{\gamma}) = 0 \text{ whenever } \gamma_i = N_i + 1 \text{ for some } i\right\}, \tag{33}$$

and

$$D^h = \left\{\overrightarrow{p} = (\overrightarrow{p_1}, \dots, \overrightarrow{p_m}, p_\gamma)\colon \Omega^h \times \widetilde{\Gamma} \to \mathbb{R}^2 \times \cdots \times \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^{2m+1}\colon |\overrightarrow{p_i}| \leq 1, (-1)^m p_\gamma \leq \rho \text{ and } \overrightarrow{p_i}(x^h, \overrightarrow{\gamma}) = 0 \text{ if } \gamma_i \neq 0\right\}. \tag{34}$$

This can be written

$$\min_{\phi^h\in C^h} \max_{\overrightarrow{p}\in D^h} \langle A\phi^h, \overrightarrow{p}\rangle_{\mathbb{R}^{2m+1}}, \tag{35}$$

where A is the linear operator that maps

$$\phi^h \mapsto (\nabla^h\phi^h, \dots, \nabla^h\phi^h, D^m_{1,\dots,m}\phi^h) \in \mathbb{R}^{2m+1}. \tag{36}$$

Following Chambolle and Pock (2010), Pock et al. (2009) and Pock et al. (2010), we use the following modified Arrow-Hurwicz-Uzawa algorithm (Popov 1980; Arrow et al. 1958); see Esser et al. (2010) for an overview of primal-dual algorithms. Choose time steps $\tau^0_\phi, \tau^0_p > 0$ such that $\tau^0_\phi\tau^0_p\|A\|^2 < 1$, where

$$\|A\| = \sup_{\phi^h\neq 0} \frac{\|A\phi^h\|}{\|\phi^h\|} \tag{37}$$

is the operator norm of A. Along the same lines as (Pock et al. 2010, Theorem 4.1), we see that

$$\|A\|^2 = \frac{4}{h_x^2} + \frac{4}{h_y^2} + \frac{4^m}{\prod_{i=1}^m h_{\gamma_i}^2}, \tag{38}$$

where h_x and h_y are the spatial step sizes of Ω^h and $h_{\gamma_1}, \dots, h_{\gamma_m}$ are the step sizes of $\widetilde{\Gamma}$. The time step requirement reduces to

$$\tau^0_\phi\tau^0_p < \frac{1}{\frac{4}{h_x^2} + \frac{4}{h_y^2} + \frac{4^m}{\prod_{i=1}^m h_{\gamma_i}^2}}.$$

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

$$\begin{cases} (\overrightarrow{p})^{n+1} = \Pi_{D^h}\big((\overrightarrow{p})^n + \tau_p^n A(\overline{\phi}^h)^n\big), \\ (\phi^h)^{n+1} = \Pi_{C^h}\big((\phi^h)^n - \tau_\phi^n A^*(\overrightarrow{p})^{n+1}\big), \\ \theta_n = 1/\sqrt{1+2\eta\tau_\phi^n}, \\ \tau_\phi^{n+1} = \theta_n \tau_\phi^n, \qquad \tau_p^{n+1} = \tau_p^n/\theta_n, \qquad \eta > 0, \\ (\overline{\phi}^h)^{n+1} = (\phi^h)^{n+1} + \theta_n((\phi^h)^{n+1} - (\phi^h)^n) \end{cases} \tag{39}$$

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,

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

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 (35) (cf. Chambolle and Pock 2010, Theorem 2). Finally, note that we have used the acceleration technique of Pock et al. (2010) for first-order primal-dual algorithms to accelerate the convergence. The acceleration is designed to be efficient only if the objective function has some smoothness (uniform convexity of one of the two terms in the saddle-point problem). Our proposed energy (32) does not hold this property. However, experiments showed that the accelerated algorithm is still 5–10% faster than a regular primal-dual algorithm.

4 Completely Convex Chan-Vese Model

In Sect. 2, we described the method from Chan et al. (2006) to find global minimizers of the Chan-Vese problem

$$\min_{\Sigma\subset\Omega} MS(\Sigma,c_1,c_2) \tag{40}$$

where $c_1, c_2 \in \mathbb{R}$ are fixed. Now, using the vector-valued minimization technique developed in Sect. 3, we propose a method to compute tight approximations of the optimal solutions of the Chan-Vese problem when the values c_1, c_2 are no longer fixed.

The idea is that the constants $c_1, c_2 \in \mathbb{R}$ can be represented by functions $v_1, v_2 \colon \Omega \to \mathbb{R}$ provided $\nabla v_1 = \nabla v_2 = 0$. Suppose the image is normalized so that $I\colon \Omega \to [0,1]$. Since our framework requires each component of the unknown vector-valued function to take values in a discrete set, we choose a discretization of the unit interval with step size $h = \frac{1}{N_v}$, yielding the minimization problem

$$\min_{u\colon \Omega\to\{0,1\},v_1,v_2\colon\Omega\to\{0,1,\ldots,N_v\}} \bigg\{ E_{\mathrm{BCB}}(u,v_1,v_2) = \int_\Omega |\nabla u| + \int_\Omega u(hv_1 - I)^2 + (1-u)(hv_2-I)^2\,dx \bigg\}$$
$$\text{subject to} \quad \nabla v_1 = \nabla v_2 = 0. \tag{41}$$

To enforce the constraints, we use method similar to the augmented Lagrangian method used in Brown et al. (2009). In this case, we set all the multipliers equal to zero and thus the method is simply a penalty method. More precisely, to solve the minimization problem

$$\min_{\overrightarrow{u}\colon\Omega\to\Gamma} F(\overrightarrow{u}) \quad \text{subject to} \quad H(\overrightarrow{u}) = 0, \tag{42}$$

we solve the sequence of minimization problems

$$\min_{\overrightarrow{u}\colon\Omega\to\Gamma} F_j(\overrightarrow{u}) := F(\overrightarrow{u}) + r_j \left\|H(\overrightarrow{u})\right\|^2_{L^2(\Omega)} \tag{43}$$

for $j = 1, 2, \ldots$. Suppose that the penalty parameters are such that $0 < r_1 < r_2 < \cdots$ with $r_j \to \infty$. If $\overrightarrow{u}_j$ is a global minimizer of F_j, then any limit point $\overrightarrow{u}^*$ of $\{\overrightarrow{u}_j\}_{j=1}^\infty$ is a global solution of F subject to the constraint $H = 0$.

Let $\overrightarrow{u} = (u, v_1, v_2)$. Observe that $\nabla v_1 = \nabla v_2 = 0$ if and only if

$$H(\overrightarrow{u}) := (|\nabla v_1| + |\nabla v_2|)^{1/2} = 0, \tag{44}$$

and the L^2 norm of $H(\overrightarrow{u})$ is simply the sum of the total variations of v_1 and v_2, i.e.,

$$\left\|H(\overrightarrow{u})\right\|^2_{L^2(\Omega)} = \int_\Omega |\nabla v_1| + \int_\Omega |\nabla v_2|. \tag{45}$$

Hence, we have reduced the global Chan-Vese problem (1) to a sequence of minimization problems

$$\min_{u,v_1,v_2} \bigg\{ \int_\Omega |\nabla u| + u(hv_1-I)^2 + (1-u)(hv_2-I)^2 + r_j|\nabla v_1| + r_j|\nabla v_2| \bigg\}, \tag{46}$$

where the r_j are strictly increasing and tend to infinity. The minimization is taken over $u\colon \Omega \to \{0,1\}$ and $v_1, v_2\colon \Omega \to \{0,1,\ldots,N_v\}$. For fixed j, this minimization problem is in the class of non-convex problems that can be solved using the general framework of Sect. 3. We have

$$\begin{aligned}\overrightarrow{u} &= (u_1, u_2, u_3)\\ &= (u, v_1, v_2)\colon \Omega \to \Gamma := \{0,1\} \times \{0,1,\dots,N_v\}^2 \qquad (47)\end{aligned}$$

and

$$\begin{aligned}E_r(\overrightarrow{u}) = \int_\Omega |\nabla u_1| + r\int_\Omega |\nabla u_2| + r\int_\Omega |\nabla u_3|\\ + \int_\Omega \rho(x, u_1, u_2, u_3)\,dx \qquad (48)\end{aligned}$$

where

$$\rho(x, u_1, u_2, u_3) = u_1(hu_2 - I(x))^2 + (1-u_1)(hu_3 - I(x))^2. \qquad (49)$$

Using the method of Sect. 3, we obtain the minimization problem $\min_{\phi\in X} F_r(\phi)$, where

$$\begin{aligned}F_r(\phi) = \int_\Omega \left|\nabla\phi(x, \overrightarrow{e}_1)\right| + r\sum_{\ell=1}^{N_v}\sum_{i=2}^{3}\int_\Omega \left|\nabla\phi(x, \ell\overrightarrow{e}_i)\right|\\ - \int_\Omega \sum_{\overrightarrow{\gamma}\in\Gamma} \rho(x, \overrightarrow{\gamma}) D^3\phi\,dx \qquad (50)\end{aligned}$$

and X is the set defined in (25). This leads to our algorithm for globally solving the Chan-Vese problem, shown in Algorithm 2.

Algorithm 2 Completely Convex Chan-Vese

Initialize variables {Algorithm is independent of initialization}
while not converged **do**
  $\phi^* = \operatorname{argmin}_{\phi\in X} F_r(\phi)$
  increase r
end while
$(u, c_1, c_2) \leftarrow \text{vertex}(1_{\{\phi^*\geq 1\}})$

The final step of the algorithm indicates that our solution is given by the principal vertex of the function $1_{\{\phi^*\geq 1\}}$, which is guaranteed to be a box function by Lemma 6. Due to Theorem 7, we obtain a global minimizer whenever the certificate condition

$$F(\phi^*) = F(1_{\{\phi^*\geq 1\}}) \qquad (51)$$

holds.

When the certificate condition is approximately satisfied, i.e. when the values of the functions $F(\phi^*)$ and $F(1_{\{\phi^*\geq 1\}})$ are close, we can make precise statements about the approximation of the obtained solution to the global minimizer of the original problem.

Proposition 8 *Let $\epsilon > 0$. Suppose that for $j = 1, 2, \dots,$ we have the bound*

$$\left|F_{r_j}(\phi_j^*) - F_{r_j}(1_{\{\phi_j^*\geq 1\}})\right| < \epsilon, \qquad (52)$$

where ϕ_j^* is the global minimizer of F_{r_j} over $\phi \in X$. Then the sequence of functions

$$\begin{aligned}u_j &:= \int_{\gamma_1} 1_{\{\phi_j^*\geq 1\}}(x, \gamma_1, \gamma_2 = 0, \gamma_3 = 0)\,d\gamma_1,\\ v_{1j} &:= \int_{\gamma_2} 1_{\{\phi_j^*\geq 1\}}(x, \gamma_1 = 0, \gamma_2, \gamma_3 = 0)\,d\gamma_2,\\ v_{2j} &:= \int_{\gamma_3} 1_{\{\phi_j^*\geq 1\}}(x, \gamma_1 = 0, \gamma_2 = 0, \gamma_3)\,d\gamma_3\end{aligned}$$

contains a subsequence that converges to functions $\tilde{u}, \tilde{v_1}, \tilde{v_2}$ such that $\nabla\tilde{v_1} = \nabla\tilde{v_2} = 0$ and

$$\left|E_{\text{BCB}}(\tilde{u}, \tilde{v_1}, \tilde{v_2}) - E^*\right| < \epsilon,$$

where E^* is the optimal value of E_{BCB} or equivalently energy MS in (1).

Proof The proof is given in the appendix and is mainly based on Lemma 7 (also provided in the Appendix). □

5 Numerical Results

5.1 Original Chan-Vese Model

This section illustrates the problem of non-convexity of the original Chan-Vese method (Chan and Vese 2001). The alternating Chan-Vese algorithm based on Chan et al. (2006) is given in Algorithm 1. The inputs $c_1^{\text{init}}, c_2^{\text{init}}$ are computed from an initial function u_{init} using (6). The output function u_{final} is thresholded at 0.5, which provides a binary function $u^\diamond := 1_{\{u_{\text{final}}\geq 0.5\}}$. It is clear on Figs. 4 and 5 that the choice of the initial function u_{init} leads to different segmentation results with different energy values. This shows that the alternating Chan-Vese method is not guaranteed to provide global solutions of the original non-convex minimization problem (1).

5.2 Completely Convex Chan-Vese Model

In this section, we illustrate our segmentation method on several images taken from the Berkeley and Weizmann data sets (Berkeley Dataset; Weizmann Dataset), see Figs. 6–16. We remind that the proposed relaxation method is guaranteed to compute a global minimizer of the Chan-Vese model under certain conditions, which are represented by the F-certificate in Theorem 7, i.e. $F(\phi^*) = F(1_{\{\phi^*\geq 1\}})$. If the F-certificate does not hold exactly but is still approximately

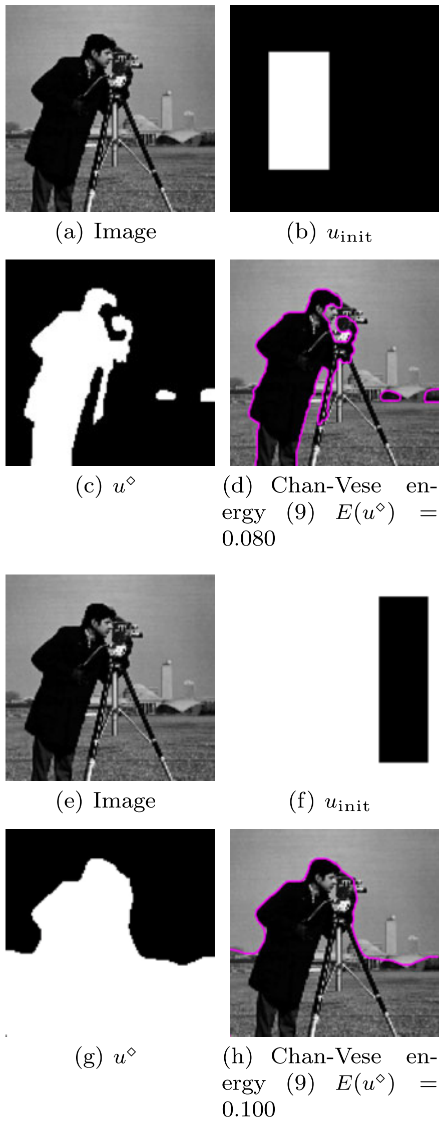


Fig. 4 Standard alternating Chan-Vese segmentation method (Chan and Vese 2001). *Rows 1 and 2* present two segmentation results for two different initial functions *u*

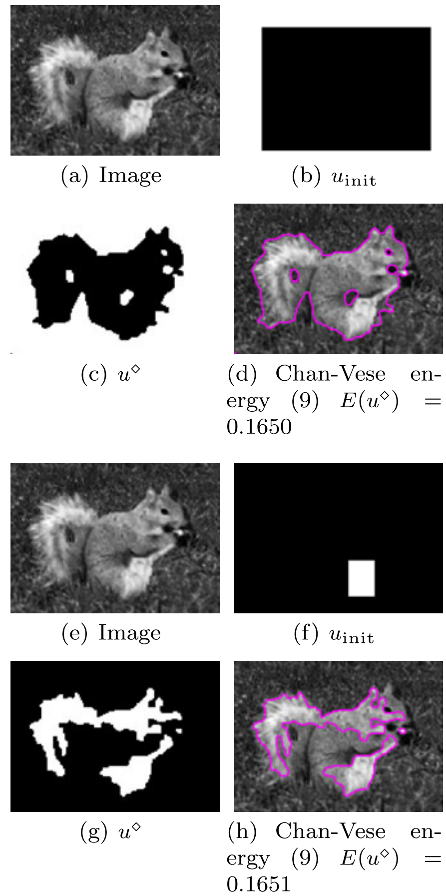


Fig. 5 Standard alternating Chan-Vese segmentation method (Chan and Vese 2001). *Rows 1 and 2* present two segmentation results for two different initial functions *u*

satisfied, i.e. $F(\phi^*)$ being close to $F(1_{\{\phi^*\geq 1\}})$, then Proposition 8 states that the thresholded solution is close (in terms of the value of the objective function) to the optimal solution of the original non-convex problem. Therefore, we will check the values of energies $F(\phi^*)$ and $F(1_{\{\phi^*\geq 1\}})$ in all experiments. The closer these two values are the closer the relaxed solution is to the optimal solution.

The first experiment we considered is the *cameraman* image on Fig. 6(a). Using the algorithm proposed in Sect. 3.2, we obtained $\phi^\star$, the non-thresholded function $u^\dagger(x) := \int_{\gamma_1} \phi^\star(x, \gamma_1, \gamma_2 = 0, \gamma_3 = 0)d\gamma_1$ on Fig. 6(b) and the thresholded/binary segmentation function $u^\star(x) := \int_{\gamma_1} 1_{\{\phi^\star\geq 1\}}(x, \gamma_1, \gamma_2 = 0, \gamma_3 = 0)d\gamma_1$ on Fig. 6(c). The solution $1_{\{\phi^\star\geq 1\}}$ is a global solution of the original Chan-Vese problem (1) if the F-certificate, which is a sufficient condition of optimality, holds. For the *cameraman* experiment, the F-certificate does not hold exactly but relatively well as we have $F(\phi^*) = 27.08$ and $F(1_{\{\phi^*\geq 1\}}) = 27.59$, i.e. 1.8% of relative difference between the two energy values. Note that the first coordinate of the vertex $1_{\{\phi^*\geq 1\}}$ is the minimizer u, which corresponds to the two-phase segmentation and is shown in Fig. 6(b). The second and third coordinates of the vertex are the minimizers v_1 and v_2, which are constants due to the constraint enforced by the penalty term. The value of the Chan-Vese energy is computed using (9) and (6) for c_1, c_2.

Figures 7–13 present other segmentation results for which the F-certificate also holds relatively well meaning that the computed solutions are tight approximations of the global solutions of the original Chan-Vese problem. Tables 1 and 2 provide the memory allocation (i.e. image size) and the running time for the different segmentation results.

Next, we consider the extension of the two-phase Chan-Vese segmentation model to *three phases*, see Fig. 14(a). We

Table 1 Memory allocation (i.e. image size) and running time for the original (*two-phase*) Chan-Vese segmentation model

Image	Cameraman	Squirrel	Boat	Leaf
Size $N_x \times N_y \times N_u \times N_v^2$	$128 \times 127 \times 2 \times 5^2$	$144 \times 105 \times 2 \times 5^2$	$128 \times 187 \times 2 \times 5^2$	$128 \times 87 \times 2 \times 5^2$
$F(\phi^*)$	27.08	17.07	12.98	14.884
$F(1_{\{\phi^* \geq 1\}})$	27.59	17.15	13.15	14.886
Relative difference	1.8%	0.4%	1.2%	0.01%
# iterations to converge	9000	14500	7000	9500
CPU (sec)	950	1711	1092	684

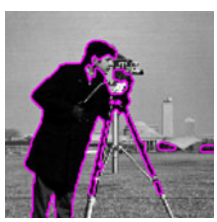
Table 2 Memory allocation (i.e. image size) and running time for the original (*two-phase*) Chan-Vese segmentation model

Image	Lung	Plane	Banzai	Planes
Size $N_x \times N_y \times N_u \times N_v^2$	$128 \times 106 \times 2 \times 5^2$	$192 \times 128 \times 2 \times 5^2$	$191 \times 128 \times 2 \times 5^2$	$190 \times 126 \times 2 \times 5^2$
$F(\phi^*)$	21.55	11.65	23.01	25.63
$F(1_{\{\phi^* \geq 1\}})$	21.56	11.66	23.32	25.66
Relative difference	0.04%	0.08%	1.3%	0.1%
# iterations to converge	2000	16000	9500	7000
CPU (sec)	172	3008	1482	1260

(a) Image

(b) $u^\dagger$ (non-thresholded solution)

(c) $u^\star$ (thresholded solution)



(d) Chan-Vese energy (9) $E(u^\star) = 0.066$

Fig. 6 Our segmentation method provides a tight approximation of the original Chan-Vese optimal solution (1) because the F-certificate of optimality holds relatively well as energies $F(\phi^*) = 27.08$ and $F(1_{\{\phi^* \geq 1\}}) = 27.59$ have 1.8% of relative difference

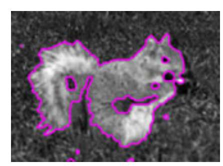
consider the natural extension of (1), that is

$$\min_{\Sigma_1, \Sigma_2, \Sigma_3, c_1, c_2, c_3} \sum_{i=1}^{3} \mathrm{Per}(\Sigma_i) + \int_{\Sigma_i} (c_i - I(x))^2 \, dx$$

(a) Image

(b) $u^\dagger$ (non-thresholded solution)

(c) $u^\star$ (thresholded solution)



(d) Chan-Vese energy (9) $E(u^\star) = 0.0862$

Fig. 7 Our segmentation method provides a tight approximation of the original Chan-Vese optimal solution (1) because the F-certificate of optimality holds relatively well as energies $F(\phi^*) = 17.07$ and $F(1_{\{\phi^* \geq 1\}}) = 17.15$ have 0.4% of relative difference

$$\text{s.t.} \quad \bigcup_{i=1}^{3} \Sigma_i = \Omega, \quad \Sigma_i \cap \Sigma_j = \emptyset \quad \forall i \neq j. \tag{53}$$

A good relaxation problem of (53) is given by Lie et al. (2006):

$$\min_{u \in \{1,2,3\}, c_1, c_2, c_3} \int_\Omega |\nabla u| + \sum_{i=1}^{3} \psi_i(u)(c_i - I)^2, \tag{54}$$

(a) Image (b) $u^\star$ (c) Segmentation

Fig. 8 Tight approximation of the original Chan-Vese optimal solution as $F(\phi^*) = 12.98$ and $F(1_{\{\phi^*\geq 1\}}) = 13.15$, i.e. 1.2% of relative difference

(a) Image (b) $u^\star$ (c) Segmentation

Fig. 9 Tight approximation of the original Chan-Vese optimal solution as $F(\phi^*) = 14.884$ and $F(1_{\{\phi^*\geq 1\}}) = 14.886$, i.e. 0.01% of relative difference

(a) Image (b) $u^\star$ (c) Segmentation

Fig. 10 Tight approximation of the original Chan-Vese optimal solution as $F(\phi^*) = 21.55$ and $F(1_{\{\phi^*\geq 1\}}) = 21.56$, i.e. 0.04% of relative difference

(a) Image (b) $u^\star$ (c) Segmentation

Fig. 11 Tight approximation of the original Chan-Vese optimal solution as $F(\phi^*) = 11.65$ and $F(1_{\{\phi^*\geq 1\}}) = 11.66$, i.e. 0.08% of relative difference

where ψ_i are indicator functions of the region $\{u = i\}$. Here $\psi_1 = \frac{1}{2}(u-2)(u-3)$, $\psi_2 = (1-u)(u-3)$, and $\psi_3 = \frac{1}{2}(u-1)(u-2)$. The relaxed problem (54) is equivalent to (53) if and only if the zero-level sets do not overlap nontrivially, see Brown et al. (2010). We now consider the straightforward extension of (41) to three phases:

(a) Image (b) $u^\star$ (c) Segmentation

Fig. 12 Tight approximation of the original Chan-Vese optimal solution as $F(\phi^*) = 23.01$ and $F(1_{\{\phi^*\geq 1\}}) = 23.32$, i.e. 1.3% of relative difference

(a) Image (b) $u^\star$ (c) Segmentation

Fig. 13 Tight approximation of the original Chan-Vese optimal solution as $F(\phi^*) = 25.63$ and $F(1_{\{\phi^*\geq 1\}}) = 25.66$, i.e. 0.1% of relative difference

$$\min_{u\in\{1,2,3\},v_1,v_2,v_3\in\{0,1,\dots,N_v\}} \left\{ \int_\Omega |\nabla u| + \sum_{i=1}^{3} \psi_i(u)(hv_i - I)^2\, dx \right\} \tag{55}$$

subject to $\quad \nabla v_i = 0 \quad \forall i = \{1, 2, 3\}.$

Using the method of Sect. 3, we obtain a convex minimization problem $\min_{\phi\in X} F_r(\phi)$ to solve (55). Figure 14(b) presents the segmentation/binary function $u^\star(x) := \int_{\gamma_1} 1_{\{\phi^\star\geq 1\}}(x, \gamma_1, \gamma_2 = 0, \gamma_3 = 0, \gamma_4 = 0)d\gamma_1$. Again, the solution $1_{\{\phi^\star\geq 1\}}$ is a global solution of the three-phase Chan-Vese problem if the F-certificate holds. For the *brain* experiment, the F-certificate holds relatively well as we have $F(\phi^*) = 19.46$ and $F(1_{\{\phi^*\geq 1\}}) = 19.47$, i.e. 0.05% of relative difference between the two energy values. Note that the first coordinate of the vertex $1_{\{\phi^*\geq 1\}}$ is the minimizer u, which corresponds to the three-phase segmentation and shown on Fig. 14(b). The second, third and fourth coordinates of the vertex are the minimizers v_1, v_2, and v_3 which are constants due to the constraint enforced by the penalty term.

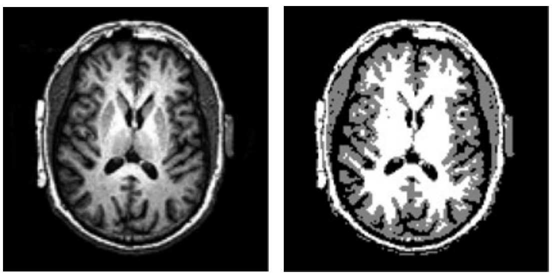
Figures 15 and 16 present other segmentation results, for which the proposed solutions are also tight approximations of the optimal solutions of the three-phase Chan-Vese problem. Table 3 provides the memory allocation (i.e. image size) and the running time for the different segmentation results.

5.3 Discussions

In this section, we provide some details and observations regarding the proposed algorithm.

Table 3 Memory allocation (i.e. image size) and running time for the *three-phase* Chan-Vese segmentation model

Image	Brain	Landscape	Swan
Size $N_x \times N_y \times N_u \times N_v^3$	$128 \times 128 \times 3 \times 4^3$	$192 \times 128 \times 3 \times 4^3$	$128 \times 85 \times 3 \times 5^3$
$F(\phi^*)$	19.90	37.84	18.673
$F(1_{\{\phi^* \geq 1\}})$	19.91	37.85	18.675
Relative difference	0.05%	0.02%	0.01%
# iterations to converge	5000	2000	30000
CPU (sec)	6890	16000	25800

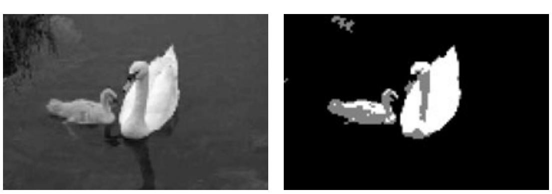


(a) Image (b) $u^\star$

Fig. 14 Our segmentation method provides a tight approximation of the *three-phase* Chan-Vese problem because the F-certificate of optimality holds well as energies $F(\phi^*) = 19.90$ and $F(1_{\{\phi^* \geq 1\}}) = 19.91$ have 0.05% of relative difference

(a) Image (b) $u^\star$

Fig. 15 Tight approximation of the *three-phase* Chan-Vese optimal solution as $F(\phi^*) = 37.84$ and $F(1_{\{\phi^* \geq 1\}}) = 37.85$, i.e. 0.02% of relative difference



(a) Image (b) $u^\star$

Fig. 16 Tight approximation of the *three-phase* Chan-Vese optimal solution as $F(\phi^*) = 18.673$ and $F(1_{\{\phi^* \geq 1\}}) = 18.675$, i.e. 0.01% of relative difference

The convergence of the minimization algorithm is checked with two standard stopping conditions. If the L^2 difference between two functions s.a. $\|\phi^{k+500} - \phi^k\|_2$, ϕ^{k+500} being the value of ϕ^k after 500 iterations, is below an arbitrary threshold then the iterative scheme stops and if the number of iterations is larger than a maximum number (30,000) then the iterative scheme stops as well.

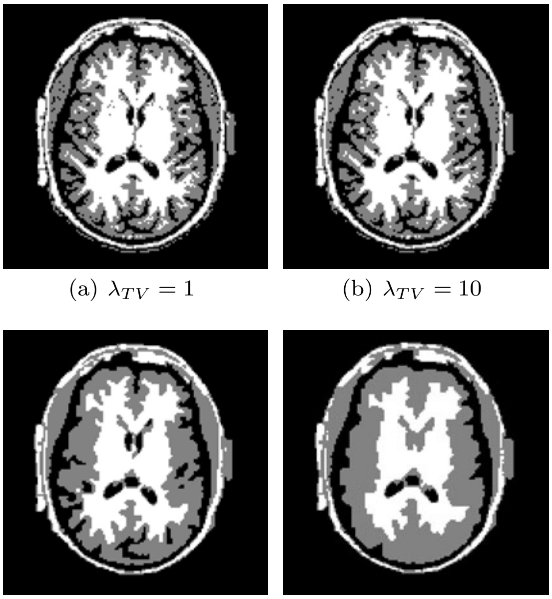
We observed in the experiments that one single iteration of the augmented Lagrangian-based scheme in Algorithm 2 is actually enough to satisfy the constraints $\nabla v_i = 0$ (after thresholding) at each pixel, which corresponds to constant functions representing the constants c_i. More precisely, a single (fixed) value of r large enough ($r = 1{,}000$ or $r = 10{,}000$) can satisfy the constraints $\nabla v_i = 0$ at each pixel. Note that when the constraints $\nabla v_i = 0$ hold, then increasing the value r does not change the minimizing solution. Indeed, increasing the value of r does not change the energy value of $F(\phi)$ because the penalty term $r \int |\nabla v_i|$ is zero $\forall r$ when the constraints $\nabla v_i = 0$ are satisfied.

We also observed that if the algorithm is run for a long time then the minimizer $\phi^\star$ will be equal to 1 for all pixels. However, the computational time to reach the value 1 is too large, so we decided to threshold $\phi^\star$ at 0.99 in all experiments to save time. Thresholding at 0.99 instead of 1 may affect Theorem 7 and Proposition 8 but experiments showed that the solutions thresholding at 0.99 are still close to the optimal solutions w.r.t. the F-certificate.

The proposed energy to minimize is convex but not strictly convex, which means that there may be more than one global minimizer. However, experiments show that the algorithm is attracted by one global minimizer.

The speed and the memory allocation of the proposed segmentation method can be an issue. The memory allocation is $O(N_x \cdot N_y \cdot N_u \cdot N_v^2)$ for the two-phase segmentation model and $O(N_x \cdot N_y \cdot N_u \cdot N_v^3)$ for the three-phase model. N_x, N_y are the size in pixel units of the image. The function u is discretized with 2 or 3 values depending the number of phases, so $N_u \in \{2, 3\}$. The problem is the number N_v. How many discretization levels do we need for the mean values c_i? The more the better but we cannot choose a large N_v value otherwise the memory allocation of our algorithm is too large. Therefore we decided to take $N_v = 5$ for the two-phase problem and $N_v = 4$ for the three-phase problem, to allow reasonable memory allocation and processing time. Improving this part is part of our future work.

We tested different degrees of smoothness on Fig. 17 and Table 4. For small and medium values of the smooth-



(a) $\lambda_{TV} = 1$ (b) $\lambda_{TV} = 10$

(c) $\lambda_{TV} = 100$ (d) $\lambda_{TV} = 200$

Fig. 17 Segmentation results for different regularization parameters

Table 4 Values of energies for different regularization parameters

λ_{TV}	$F(\phi^*)$	$F(1_{\{\phi^* \geq 1\}})$
1	19.805	19.807
10	19.90	19.91
100	23.55	23.67
200	28.78	29.71

ness parameter then the F-certificate holds rather well. For large values of the smoothness parameter then the difference between $F(\phi^\star)$ and $F(1_{\{\phi^\star \geq 1\}})$ increases. We do not know exactly why. It may be related with the running time of the algorithm. It is indeed known that large smoothness parameters can increase significantly the running time of TV-based algorithms. Therefore the F-certificate may be not satisfied because the algorithm may have not converged. It can be also intrinsic to our method, i.e. the difference between $F(\phi^\star)$ and $F(1_{\{\phi^\star \geq 1\}})$ increases when the smoothness parameter increases. Fortunately, we have noticed that the relative difference between the two energy values $F(\phi^\star)$ and $F(1_{\{\phi^\star \geq 1\}})$ does not vary much, between 0.01–1.8%, for images borrowed from the Berkeley and the Weizmann data sets.

Finally, observe that the discretization using the ℓ^2 norm may lead to a finite-dimensional problem where the (generalized) coarea formula does not hold for the discretized energy. Therefore, most of the lemma/theorems do not apply directly to the discretized energy, although experiments agreed with the continuous theory developed in this paper.

6 Concluding Remarks

In this paper, we have considered the Chan-Vese problem

$$\min_{\Sigma \subset \Omega, c_1, c_2 \in \mathbb{R}} MS(\Sigma, c_1, c_2), \tag{56}$$

and its multiphase version in (53). In the original work of Chan and Vese (2001), the authors developed a novel level-set approach to efficiently compute solutions. Later, Chan et al. (2006) introduced a convex relaxation method that guarantees global solutions to the problem when c_1 and c_2 are fixed. This yields an alternating scheme between the global solution and updating c_1 and c_2. We have proposed a completely convex method to solve this problem. The method is guaranteed to compute a global minimizer under certain conditions. These conditions do not hold exactly in practice, so we cannot guarantee to find exact global minimizers of the original Chan-Vese problem. However, our experiments show that our method computes tight approximations of the optimal solutions as the error on the certificate of optimality is only 0.01–1.8% for all experiments.

We can view the problem considered in this paper as the K-means problem (MacQueen 1967), which is a special case of maximum likelihood estimation (MLE). In particular, consider the case of a mixture model of two distributions of the same family with different unknown parameters. We have observed data $x = x_1, \ldots, x_n$ that is assumed to be independently drawn from the distributions and latent variables $z = z_1, \ldots, z_n$ determine the distribution from which each datum originates. The aim is to estimate the unknown parameters θ of the distributions and the mixture of the data. In other words, we seek to determine θ and z that maximize the likelihood of the observed data. Since this optimization problem is often intractable, other approaches, such as the expectation-maximization (EM) algorithm, are often used. EM is an iterative method that alternates between computing the expectation (E) of the likelihood using the current estimate of the latent variables, and maximizing (M) the expected value with respect to the unknown parameters, which then determines the distribution of the latent variables in the next iteration. To see the relationship with our setting, consider the observed data x as the image I, the latent variables z as the labeling Σ (i.e., the binary labeling function u), and the unknown parameters θ as the optimal constants c_1 and c_2. Continuing with the analogy, we can view Algorithm 1 as an EM algorithm. The E step is finding the segmentation by globally solving the convex optimization problem when the constants are fixed; the M step is the straightforward calculation of updating the constants for the given segmentation. Furthermore, solving the overall MLE problem is analogous to solving the global Chan-Vese problem. We leave for future work the task of attempting to generalize our methods to global optimization problems similar to those in the MLE setting.

It was recently introduced in Pock et al. (2009) a convex relaxation technique based on the calibration theory (Alberti et al. 1999) that computes an approximate global solution to the original Mumford-Shah energy, i.e. the piecewise smooth case. Mumford and Shah (1989) showed that the piecewise constant energy (the one considered in this paper) is the natural limit energy of the piecewise smooth energy when the data fidelity parameter tends to zero (equivalently when the regularization parameter goes to infinity). So in this sense, our proposed approach can be considered as a limiting case of Pock et al. (2009). It would be an interesting work to study the limit case of the algorithm proposed in Pock et al. (2009) and see what discrete continuation principle (that increases the regularization parameter to infinity) may produce the same solutions.

It was also recently proposed in Strandmark et al. (2009) another method to compute global solutions to the two-phase Chan-Vese segmentation problem. The proposed method is based on the convex relaxation method (Chan et al. 2006) and thresholding techniques. Their overall algorithm performs most likely faster than our segmentation technique. However, our segmentation method is more general than Strandmark et al. (2009) because it is not restricted to the two-phase segmentation problem. Our framework was originally developed to solve the multi-phase segmentation problem (Brown et al. 2010). Therefore, in order to illustrate the generality of our framework, we considered solutions to the three-phase segmentation problem. The proposed preliminary three-phase segmentation algorithm is rather slow and memory consuming. Nevertheless, it was important to show the generality of our segmentation framework to compute solutions to more than two phases.

Future work will focus on improving the computational efficiency of the proposed method. Indeed, although the proposed method is the first completely convex relaxation method for the 2-phase and multi-phase Chan-Vese model, it is still limited in practice as the number of discretized values for the feature c_1, c_2 is only $N_v = 5$. However, new techniques s.a. (Shekhovtsov et al. 2008; Goldluecke and Cremers 2010) are promising to increase significantly the number of discretized values for c_1, c_2.

Acknowledgements The authors would like to thank Tom Goldstein for very helpful discussion, and the referees and the associate editors for their valuable general and technical comments that helped to improve the quality of this paper. Xavier Bresson is supported by the Hong Kong RGC under Grant GRF110311.

Appendix

Lemma 2 *Let F be defined by* (23). *Then F satisfies the generalized coarea formula* (24).

Proof Recall that the coarea formula (Fleming and Rishel 1960) for functions g of bounded variation states that

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

Using some elementary facts about total variation, namely that

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

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

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

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

$$\begin{aligned}&\sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \int_\Omega \left|\nabla \phi(x, \ell \vec{e_i})\right| \\ &\quad = \int_0^1 \sum_{i=1}^{m} \sum_{\ell=1}^{N_i} \int_\Omega \left|\nabla 1_{\{\phi(x,\ell\vec{e_i})\geq t\}}\right| dt.\end{aligned}$$

We also have $\phi(x, \vec{\gamma}) = \int_0^{\phi(x,\vec{\gamma})} dt = \int_0^1 1_{\{\phi\geq t\}}\, dt$, often referred to as the layer-cake formula (Lieb and Loss 2001, pp. 26–27). Thus, by linearity,

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

and (24) follows. □

Lemma 3 *Let Y be any subset of functions $\phi: \Omega \times \widetilde{\Gamma} \to [0,1]$ 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 almost 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 $F(\phi^*) \leq F(\phi')$. By minimality, $F(\phi^*) \leq F(1_{\{\phi^*\geq t\}})$, and so from (24) 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 $t \in (0, 1]$. □

Lemma 4 *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 (22) for all $t \in (0, 1]$.*

Proof We apply Lemma 3 to the minimizer ϕ^* of F over the set $Y = X$. By assumption, $1_{\phi^* \geq t}$ is a box function for all $t \in (0, 1]$, and thus, since box functions are a subset of X, we have $1_{\phi^* \geq t}$ in X for all $t \in (0, 1]$. Applying the lemma, it follows that $1_{\phi^* \geq t}$ is a minimizer of F over Y, and thus is a minimizer of F over the set of all box functions, i.e., $1_{\phi^* \geq t}$ is a minimizer of (22) for all $t \in (0, 1]$. □

Lemma 5 *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 (22).*

Proof By assumption, ϕ^* is a minimizer of F over X, $F(\phi^*) = F(1_{\phi^* \geq t})$, and $1_{\phi^* \geq t}$ is a box function. Since box functions are a subset of X, we see that $1_{\phi^* \geq t}$ is a minimizer of F over X, and again by the subset property $1_{\phi^* \geq t}$ must be a minimizer of F over box functions, i.e., $1_{\phi^* \geq t}$ is a minimizer of (22). □

Lemma 6 *For all $x \in \Omega$, suppose that $\phi\colon \Omega \times \widetilde{\Gamma} \to [0, 1]$ lies in the set X defined in (25), 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

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

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

Proof We split up the proof into two parts. First, we show that (26) 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 (26) implies

$$(-1)^k D^k_S \phi \geq 0 \quad \text{for all } 1 \leq k \leq m \tag{57}$$

and $S \subset \{1, \ldots, m\}$ with $|S| = k$. Inductively, and by relabeling indices if necessary, it suffices to show that (26) 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

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

Applying this recursively, we conclude that there exists $\overrightarrow{\gamma}' = (\gamma_1', \ldots, \gamma_m')$ with $\gamma_m' = N_m + 1$ such that $(-1)^{m-1} \times D^{m-1}_{1,\ldots,m-1} \phi(x, \overrightarrow{\gamma}') < 0$. 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 Fig. 18.

Recall the boundary conditions required by (a) and (b); see Fig. 18(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 Fig. 18(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

$$\begin{aligned} &\phi_x(\overrightarrow{\gamma} + \overrightarrow{e_i}) = 1 \quad \text{and} \quad \phi_x(\overrightarrow{\gamma} + \overrightarrow{e_j}) = 1 \\ &\implies \quad \phi_x(\overrightarrow{\gamma} + \overrightarrow{e_i} + \overrightarrow{e}_j) = 1 \end{aligned} \tag{58}$$

for all $\overrightarrow{\gamma}$ and indices $i \neq j$ in which these expressions are defined. Note that the left-hand side implicitly means that

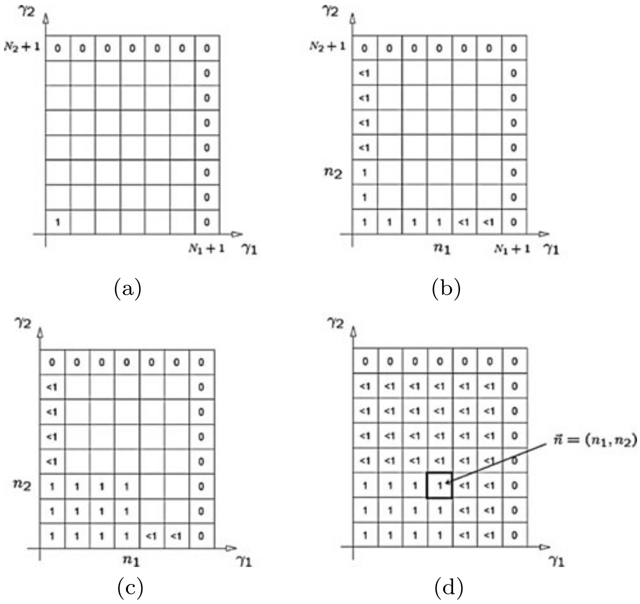


Fig. 18 An illustration of the proof of Lemma 6. See the proof for the explanations of each of the steps (**a**)–(**d**). In particular, (**d**) shows $\vec{n}$, which we see is the principal vertex of $1_{\{\phi\geq 1\}}$

$\phi_x(\overrightarrow{\gamma}) = 1$, 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 (58) forces the remaining interior entries to be subsequently filled with 1's until we form the "box" of 1's. See Fig. 18(c).

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 (58), 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., $\overrightarrow{\gamma} \npreceq \overrightarrow{n}$), 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 Fig. 18(d). We conclude that $1_{\{\phi\geq 1\}}$ is a box function with principal vertex $\overrightarrow{n}$, and the proof is complete. □

The following lemma is used to prove Proposition 8.

Lemma 7 *Let*

$$E_{r_j}(u, v_1, v_2) = E_{\mathrm{BCB}}(u, v_1, v_2) + r_j \int_\Omega |\nabla v_1| + r_j \int_\Omega |\nabla v_2|, \tag{59}$$

with $0 < r_1 < r_2 < \cdots$ *with* $r_j \to \infty$. *Any sequence* $\{u_j, v_{1j}, v_{2j}\}$ *of global minimizers of* E_{r_j} *contains a subsequence that converges in* L^1 *(and pointwise almost everywhere) to a function* $\overline{u}, \overline{v_1}, \overline{v_2} \in BV(\Omega; [0, 1])$ *that is a global solution of*

$$\min_{u,v_1,v_2\in BV(\Omega;[0,1])} E_{\mathrm{BCB}}(u, v_1, v_2) = \int_\Omega |\nabla u| + u(hv_1 - I)^2 + (1-u)(hv_2 - I)^2$$

$$\text{s.t.} \quad \nabla v_1 = \nabla v_2 = 0 \tag{60}$$

and consequently

$$\min_{u\in BV(\Omega;[0,1]),c_1,c_2\in[0,1]} E_{\mathrm{CEN}}(u, c_1, c_2) = \int_\Omega |\nabla u| + u(c_1 - I)^2 + (1-u)(c_2 - I)^2, \tag{61}$$

assuming without loss of generality that $I \in [0, 1]$ *and so are the average values* c_1, c_2.

Proof Our argument is based on the proof of (Bertsekas 1982, Prop. 2.1), which establishes the result in the finite dimensional setting.

By assumption, each function u_j, v_{1j}, v_{2j} is a global minimizer of $E_{r_j}(u, v_1, v_2)$, which means that

$$E_{r_j}(u_j, v_{1j}, v_{2j}) \leq E_{r_j}(u, v_1, v_2) \quad \text{for all } u, v_1, v_2 \in BV(\Omega; [0, 1]). \tag{62}$$

Let E^* denote the optimal value of (60). Then, for any fixed $j \geq 1$,

$$\begin{aligned} E^* &= \inf\{E_{\mathrm{BCB}}(u, v_1, v_2) : u, v_1, v_2 \in BV(\Omega; [0, 1]), \\ &\quad \nabla v_1 = \nabla v_2 = 0\} \\ &= \inf\{E_{r_j}(u, v_1, v_2) : u, v_1, v_2 \in BV(\Omega; [0, 1]), \\ &\quad \nabla v_1 = \nabla v_2 = 0\}. \end{aligned}$$

Taking the infimum of (62) over $u, v_1, v_2 \in BV(\Omega; [0, 1])$ such that $\nabla v_1 = \nabla v_2 = 0$ yields

$$E_{\mathrm{BCB}}(u_j, v_{1j}, v_{2j}) + r_j \int_\Omega |\nabla v_{1j}| + r_j \int_\Omega |\nabla v_{2j}| \leq E^*. \tag{63}$$

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

$$\int_\Omega |\nabla u_j|, \quad \int_\Omega |\nabla v_{1j}|, \quad \int_\Omega |\nabla v_{2j}|$$

is uniformly bounded. Let $\rho(u_1,u_2,u_3)=u_1(hu_2-I)^2+(1-u_1)(hu_3-I)^2$. Choosing $u, v_1, v_2 \equiv 0$ in the inequality (62) yields

$$\int_\Omega |\nabla u_j| + \rho(u_j, v_{1j}, v_{2j}) + r_j \int_\Omega |\nabla v_{1j}| + r_j \int_\Omega |\nabla v_{2j}| \le \int_\Omega I^2 = M, \quad 0 < M < \infty$$

since the function ρ is bounded, which implies

$$\int_\Omega |\nabla u_j| + r_j \int_\Omega |\nabla v_{1j}| + r_j \int_\Omega |\nabla v_{2j}| \le M.$$

Therefore the sequence $\{u_j, v_{1j}, v_{2j}\}$ is uniformly bounded in $BV(\Omega;[0,1])$. Thus, by a well-known BV compactness theorem (see, e.g., Evans and Gariepy 2000, p. 176), there exists a subsequence $\{u_{n_j}, v_{1n_j}, v_{2n_j}\}$ converging in L^1 to $\overline{u}, \overline{v_1}, \overline{v_2} \in BV(\Omega;[0,1])$. Moreover, by lower semicontinuity of the BV seminorm and by passing to a subsequence if necessary, we have

$$\int_\Omega |\nabla \overline{u}| \le \lim_{j\to\infty} \int_\Omega |\nabla u_j|. \tag{64}$$

Continuity of ρ together with Lebesgue's dominated convergence theorem implies that

$$\int_\Omega \rho(u_j, v_{1j}, v_{2j}) \to \int_\Omega \rho(\overline{u}, \overline{v_1}, \overline{v_2})$$

which, together with (64), establishes that

$$E_{\mathrm{BCB}}(\overline{u}, \overline{v_1}, \overline{v_2}) \le \lim_{j\to\infty} E_{\mathrm{BCB}}(u_j, v_{1j}, v_{2j}).$$

Consequently, taking the limit superior of (63), we have

$$E_{\mathrm{BCB}}(\overline{u}, \overline{v_1}, \overline{v_2}) + \limsup_{j\to\infty} r_j \int_\Omega |\nabla v_{1j}| + \limsup_{j\to\infty} r_j \int_\Omega |\nabla v_{2j}| \le E^*. \tag{65}$$

But since $\int_\Omega |\nabla v_{1j}| \ge 0$, $\int_\Omega |\nabla v_{2j}| \ge 0$, $r_j \to \infty$, and $E^* < \infty$, we must have $\int_\Omega |\nabla v_{1j}| \to 0$, $\int_\Omega |\nabla v_{2j}| \to 0$ and $\nabla \overline{v_1} = \nabla \overline{v_2} = 0$, otherwise the left-hand side of (65) is infinite. It follows that $E_{\mathrm{BCB}}(\overline{u}, \overline{v_1}, \overline{v_2}) \le E^*$, as was to be shown. □

Proposition 8 *Let $\epsilon > 0$. Suppose that for $j = 1, 2, \dots,$ we have the bound*

$$\left| F_{r_j}(\phi_j^*) - F_{r_j}(1_{\{\phi_j^* \ge 1\}}) \right| < \epsilon,$$

where ϕ_j^* is the global minimizer of F_{r_j} over $\phi \in X$. Then the sequence of functions

$$u_j := \int_{\gamma_1} 1_{\{\phi_j^* \ge 1\}}(x, \gamma_1, \gamma_2 = 0, \gamma_3 = 0)\, d\gamma_1,$$

$$v_{1j} := \int_{\gamma_2} 1_{\{\phi_j^* \ge 1\}}(x, \gamma_1 = 0, \gamma_2, \gamma_3 = 0)\, d\gamma_2,$$

$$v_{2j} := \int_{\gamma_3} 1_{\{\phi_j^* \ge 1\}}(x, \gamma_1 = 0, \gamma_2 = 0, \gamma_3)\, d\gamma_3,$$

contains a subsequence that converges to functions $\tilde{u}, \tilde{v}_1, \tilde{v}_2$ such that $\nabla\tilde{v}_1 = \nabla\tilde{v}_2 = 0$ and

$$\left| E_{\mathrm{BCB}}(\tilde{u}, \tilde{v}_1, \tilde{v}_2) - E^* \right| < \epsilon,$$

where E^* is the optimal value of E_{BCB} or equivalently energy MS in (1).

Proof The bound (52) implies that $F_{r_j}(1_{\{\phi_j^* \ge 1\}}) < F_{r_j}(\phi_j^*) + \epsilon$. Since $1_{\{\phi_j^* \ge 1\}}$ is a box function, we may define

$$u'_j := \int_{\gamma_1} 1_{\{\phi_j^* \ge 1\}}(x, \gamma_1, \gamma_2 = 0, \gamma_3 = 0)\, d\gamma_1,$$

$$v_1{}'_j := \int_{\gamma_2} 1_{\{\phi_j^* \ge 1\}}(x, \gamma_1 = 0, \gamma_2, \gamma_3 = 0)\, d\gamma_2,$$

$$v_2{}'_j := \int_{\gamma_3} 1_{\{\phi_j^* \ge 1\}}(x, \gamma_1 = 0, \gamma_2 = 0, \gamma_3)\, d\gamma_3$$

and have $F_{r_j}(1_{\{\phi_j^* \ge 1\}}) = E_{r_j}(u'_j, v_1{}'_j, v_2{}'_j)$. Moreover, $E_{r_j}(u'_j, v_1{}'_j, v_2{}'_j) < E_{r_j}(u_j, v_{1j}, v_{2j}) + \epsilon$, where u_j, v_{1j}, v_{2j} is a global minimizer of E_{r_j}. Lemma 7 shows that any limit point of the functions $u'_j, v_1{}'_j, v_2{}'_j$, say $\tilde{u}, \tilde{v}_1, \tilde{v}_2$, satisfies the constraint $\nabla\tilde{v}_1 = \nabla\tilde{v}_2 = 0$. Since we know that $\nabla v_1^* = \nabla v_2^* = 0$ and $E^* = E_{\mathrm{BCB}}(u^*, v_1^*, v_2^*)$, the desired bound follows. □

References

Alberti, G., Bouchitté, G., & Dal Maso, G. (1999). The calibration method for the Mumford-Shah functional. *Comptes Rendus de L'Académie Des Sciences. Series 1, Mathematics*, *329*(3), 249–254.

Ambrosio, L., Fusco, N., & Pallara, D. (2000). *Functions of bounded variation and free discontinuity problems*. Oxford: Clarendon Press.

Arrow, K., Hurwicz, L., & Uzawa, H. (1958). *Stanford mathematical studies in the social sciences: Vol. II. Studies in linear and non-linear programming. With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow*. Stanford: Stanford University Press.

Bae, E., & Tai, X.-C. (2009). Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. In *International conference on scale space and variational methods in computer vision* (pp. 1–13).

Bae, E., Yuan, J., & Tai, X.-C. (2009). Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision*, *92*(1), 112–129.

Bae, E., Yuan, J., Tai, X.-C., & Boykov, Y. (2010). *A study on continuous max-flow and min-cut approaches. Part II: multiple linearly ordered labels* (UCLA CAM Report 10-62).

Berkeley Dataset. http://www.eecs.berkeley.edu/CS/vision.

Bertsekas, D. P. (1982). *Constrained optimization and Lagrange multiplier methods*. New York: Academic Press.
Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J., & Osher, S. (2007). Fast global minimization of the active contour/snake models. *Journal of Mathematical Imaging and Vision*, *28*(2), 151–167.
Brown, E. S., Chan, T. F., & Bresson, X. (2009). *Convex formulations for piecewise constant Mumford-Shah image segmentation* (UCLA CAM Report 09-66).
Brown, E. S., Chan, T. F., & Bresson, X. (2010). *A convex relaxation method for a class of vector-valued minimization problems with applications to Mumford-Shah segmentation* (UCLA CAM Report 10-43).
Chambolle, A., Cremers, D., & Pock, T. (2008). *A convex approach for computing minimal partitions* (Technical report TR-2008-05). Bonn: Dept. of Computer Science, University of Bonn.
Chambolle, A., & Pock, T. (2010). *A first-order primal-dual algorithm for convex problems with applications to imaging* (R.I. 685). CMAP, Ecole Polytechnique.
Chan, T. F., Esedoḡlu, S., & Nikolova, M. (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, *66*(5), 1632–1648.
Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on Image Processing*, *10*(2), 266–277.
El-Zehiry, N., Xu, S., Sahoo, P., & Elmaghraby, A. (2007). Graph cut optimization for the Mumford-Shah model. In *IASTED international conference on visualization, imaging and image processing* (pp. 182–187).
Esser, E., Zhang, X., & Chan, T. F. (2010). A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, *3*(4), 1015–1046.
Evans, L. C., & Gariepy, R. F. (2000). *Measure theory and fine properties of functions*. Boca Raton: CRC Press.
Federer, H. (1959). Curvature measures. *Transactions of the American Mathematical Society*, *93*(3), 418–491.
Fleming, W., & Rishel, R. (1960). An integral formula for total gradient variation. *Archiv der Mathematik*, *11*(1), 218–222.
Goldluecke, S., & Cremers, D. (2010). Convex relaxation for multilabel problems with product label spaces. In *European conference on computer vision* (pp. 225–238).
Goldstein, T., Bresson, X., & Osher, S. (2009). Geometric applications of the split Bregman method: segmentation and surface reconstruction. *Journal of Scientific Computing*, *45*(1–3), 272–293.
Goldstein, T., Bresson, X., & Osher, S. (2009). *Global minimization of Markov random fields with applications to optical flow* (UCLA CAM Report 09-77).
Ishikawa, H. (2003). Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(10), 1333–1336.
Lellmann, J., Becker, F., & Schnörr, C. (2009). Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In *International conference on computer vision* (pp. 646–653).
Lellmann, J., Kappes, J., Yuan, J., Becker, F., & Schnörr, C. (2009). Convex multi-class image labeling by simplex-constrained total variation. In *International conference on scale space and variational methods in computer vision* (pp. 150–162).
Lellmann, J., & Schnörr, C. (2010). *Continuous multiclass labeling approaches and algorithms* (Tech. Rep.). Heidelberg: University of Heidelberg.
Lie, J., Lysaker, M., & Tai, X.-C. (2006). A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE Transactions on Image Processing*, *15*(5), 1171–1181.
Lieb, E. H., & Loss, M. (2001). *Analysis*. Providence: Am. Math. Soc.
MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 281–297).
Mumford, D., & Shah, J. (1989). Optimal approximations of piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, *42*, 577–685.
Osher, S., & Sethian, J. A. (1988). Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, *79*(1), 12–49.
Pock, T., Chambolle, A., Bischof, H., & Cremers, D. (2009). An algorithm for minimizing the Mumford-Shah functional. In *IEEE conference on computer vision (ICCV)*.
Pock, T., Cremers, D., Bischof, H., & Chambolle, A. (2010). Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*, *2*(3), 1122–1145.
Pock, T., Schoenemann, T., Graber, G., Bischof, H., & Cremers, D. (2008). A convex formulation of continuous multi-label problems. In *European conference on computer vision (ECCV)* (pp. 792–805).
Popov, L. D. (1980). A modification of the Arrow-Hurwitz method of search for saddle points. *Matematičeskie Zametki*, *28*(5), 777–784, 803.
Shekhovtsov, A., Kovtun, I., & Hlavác, V. (2008). Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, *112*(1), 91–99.
Strandmark, P., Kahl, F., & Overgaard, N. C. (2009). Optimizing parametric total variation models. In *International conference on computer vision* (pp. 2240–2247).
Strang, G. (1983). Maximal flow through a domain. *Mathematical Programming*, *26*(2), 123–143.
Weizmann Dataset. http://www.weizmann.ac.il/~vision.
Yuan, J., Bae, E., Tai, X.-C., & Boykov, Y. (2010). *A study on continuous max-flow and min-cut approaches. Part I: binary labeling* (UCLA CAM Report 10-61).
Zach, C., Gallup, D., Frahm, J. M., & Niethammer, M. (2008). Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, modeling, and visualization* (pp. 243–252).