Simultaneous convex optimization of regions and region parameters in image segmentation models

Egil Bae, Jing Yuan and Xue-Cheng Tai

Abstract This work develops a convex optimization framework for image segmentation models, where both the unknown regions and parameters describing each region are part of the optimization process. Convex relaxations and optimization algorithms are proposed, which produce results that are independent from the initializations and closely approximate global minima. We focus especially on problems where the data fitting term depends on the mean or median image intensity within each region. We also develop a convex relaxation for the piecewise constant Mumford-Shah model, where additionally the number of regions is unknown. The approach is based on optimizing a convex energy potential over functions defined over a space of one higher dimension than the image domain.

1 Introduction

Image segmentation is one of the most important problems in image processing and computer vision. The task is to group the image pixels into several regions or objects based on their intensity values. Energy minimization has become an established paradigm to formulate such problem mathematically, where both data/scene consistency and the regularity of the segmentation regions are encoded in an energy potential. A major challenge is to solve the resulting NP-hard optimization problems numerically.

Egil Bae
Department of Mathematics, University of Bergen, Norway, e-mail: Egil.Bae@math.uib.no

Jing Yuan
Computer Science Department, University of Western Ontario, Canada, e-mail: cn.yuanjing@gmail.com

Xue-Cheng Tai
Department of Mathematics, University of Bergen, Norway, e-mail: tai@math.uib.no

Variational models for image segmentation

In this work, we focus on image segmentation with Potts regularity [16], which enforces region boundaries of minimal total length. We wish to partition the image domain Ω into n regions $\{\Omega_i\}_{i=1}^n$. For each point $x \in \Omega$ and each $i = 1,...,n$, define the data cost function $f_i(x)$ of assigning x to the region Ω_i. Image segmentation with Potts prior and predefined data cost functions can then be formulated as

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

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

where α is a parameter which controls the impact of the boundary regularization. The model (1) will be referred to as Potts model in this work. When $n > 2$ the optimization problem (1) in a discrete setting is NP-hard, therefore it is generally too difficult to find a global optimum. Algorithms exist that can compute good approximations [3] and in some cases exact solutions for level set representations of the problem [1]. In a continuous setting, several convex methods have recently appeared that may often lead to global solutions, or otherwise produce good approximations of global solutions [11, 22, 14, 2, 5].

An important example of the data costs in (1) is

$$f_i(x) = |I^0(x) - \mu_i|^\beta\,, \tag{3}$$

where $I^0(x)$ is the input image function and $\mu_i \in \mathbb{R}$, $i = 1,...,n$ are predefined region parameters and $\beta \geq 1$. An intuitive explanation of μ_i is the mean of the image intensity I^0 within region Ω_i in case $\beta = 2$, or the median value within Ω_i if $\beta = 1$. They are, however, unknown in advance. Therefore, the function f_i depends on the unknown segmentation region Ω_i and does not fit into the framework of (1). The most ideal model should not rely on a post-processing step to determine the parameters, instead the values μ_i should be part of the optimization process. In [6] and [13] such an image segmentation model was formulated as follows

$$\min_{\{\Omega_i\}_{i=1}^n} \min_{\{\mu_i\}_{i=1}^n \in X} \quad \sum_{i=1}^n \int_{\Omega_i} |I^0(x) - \mu_i|^\beta dx + \alpha \sum_{i=1}^n \int_{\partial\Omega_i} ds \tag{4}$$

subject to (2). The set X is typically the set of feasible gray values, which may be the real line $X = \mathbb{R}$ or a discrete quantization $X = \{\ell_1,...,\ell_L\}$. In contrast to (1) along with (3), the energy is minimized over both $\{\Omega_i\}_{i=1}^n$ and the region parameters $\{\mu_i\}_{i=1}^n$. The model (4) is often called the Chan-Vese model. If there is no regularization, i.e. $\alpha = 0$, (4) can be recognized as the "k-means" model, which is also an NP-hard problem. To solve (4), one possibility is alternative minimization with respect to Ω_i and μ_i until convergence as follows:
Find initialization $\{\mu_i^0\}_{i=1}^n$. For $k = 0,...$ until convergence

$$1.\ \{\Omega_i^{k+1}\}_{i=1}^n = \arg\min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} |I^0(x) - \mu_i^k|^\beta\, dx + \alpha \sum_{i=1}^n \int_{\partial\Omega_i} ds \ \text{ subject to (2)} \quad (5)$$

$$2.\ \{\mu_i^{k+1}\}_{i=1}^n = \arg\min_{\{\mu_i\}_{i=1}^n \in X} \sum_{i=1}^n \int_{\Omega_i^{k+1}} |I^0(x) - \mu_i|^\beta\, dx + \alpha \sum_{i=1}^n \int_{\partial\Omega_i^{k+1}} ds \quad (6)$$

Since (4) is not jointly convex, such a procedure does not in general produce a global minimum, but converges to a local optimum depending on the initialization of μ. Furthermore, there is no easy way to measure the quality of the converged result.

Closely related is the piecewise constant Mumford-Shah model [13], which can be expressed as a slight variation of (4) as (see Section 2.2)

$$\min_n \min_{\{\Omega_i\}_{i=1}^n} \min_{\{\mu_i\}_{i=1}^n \in X} \quad \sum_{i=1}^n \int_{\Omega_i} |I^0(x) - \mu_i|^\beta\, dx + \alpha \sum_{i=1}^n \int_{\partial\Omega_i} ds \quad (7)$$

subject to (2). The energy potential (7) is also optimized over the number of regions n. In spite of the seemingly higher complexity, we show the problem (7) is easier to tackle than (4) in the following sections.

The optimization problem (4) can also be extended to more general data cost functions $f_i(\xi_i, x)$, where ξ_i is some unknown parameter associated with region Ω_i.

$$\min_{\{\Omega_i\}_{i=1}^n} \min_{\{\xi_i\}_{i=1}^n} \quad \sum_{i=1}^n \int_{\Omega_i} f(\xi_i, x)\, dx + \alpha \sum_{i=1}^n \int_{\partial\Omega_i} ds \quad (8)$$
$$\text{s.t.} \quad \cup_{i=1}^n \Omega_i = \Omega\,, \quad \Omega_k \cap \Omega_l = \emptyset\,, \quad \forall k \neq l\,,$$

One example is $\xi_i = (\mu_i, \sigma_i)$, where μ_i is the mean and σ_i is the standard deviation of intensities in Ω_i. The data term can then be formulated as the log of the Gaussian distribution $f_i(\mu_i, \sigma_i, x) = \log(\frac{1}{\sqrt{2\pi}\sigma_i}\exp(-\frac{(I^0(x)-\mu_i)^2}{2\sigma_i^2}))$.

An image segmentation model based on the minimum description rule (MDL) was proposed [23] which places a direct penalty to the number of appearing regions of the Potts model (1). Recently, various algorithms have been proposed for computing global or good approximations of global minima [9, 20, 18] of the resulting optimization problems. We will also see that there is a close relationship between such a MDL based segmentation model and (4) or (8) if f_i and the penalty parameter are chosen in a particular way, as discussed in Section 4.

Instead of optimizing over the continuous set $X = \mathbb{R}$, the parameters μ_i in (4) or (7) can be optimized over a finite set of real numbers $X = \{\ell_1, ..., \ell_L\}$, where L is the number of elements in X. This is the case for digital images, where the image intensity is quantized and the set X consists of a finite number of gray values, for instance 256. When the L_1 data fitting term is applied (that is, $\beta = 1$ in (3)), we show there exists globally optimal $\{\mu_i\}_{i=1}^n$ that are also present in the input image. Therefore, optimizing over the finite set will produce an exact global optimum. This extends the result of [8] from 2 to any number of regions. A similar result can also be found in [21] for denoising problems with total variation regularization.

Contributions and previous work

Little work has been devoted to global optimization over the regions and parameters simultaneously in the image segmentation models (4) or (7). In order to optimize (4) or (7) over the finite set X, one simple, but very slow approach, is to minimize the energy in the Potts model (1) for every combination of $\{\mu_i\}_{i=1}^n \in X$, and finally select the combination of $\{\mu_i\}_{i=1}^n \in X$ which yields the lowest energy. Since there are a total of L^n such combinations when X contains L elements, a total number of L^n problems of the form (1) need to be solved. In case of two regions, it is known each subproblem can be solved exactly and globally [7], but L^2 subproblems need to be solved which would be rather slow as L becomes large.

Restricted to two regions and a finite set X, Darbon [8] developed an algorithm which solved a sequence of two region problems with fixed parameters μ_1 and μ_2, but avoided to check all L^2 combinations. The number of two region subproblems to be solved is still $O(L^2)$. In [4], Brown et al. cast a relaxation of the problem with two regions and quantized parameters as an optimization problem over a higher dimensional space. The size of the convex problem is $O(|\Omega|L^2)$, where $|\Omega|$ is the number of pixels, therefore the complexity of their algorithm is also $O(|\Omega|L^2)$. An approach based on the branch and bound method was proposed for two region problems in [12]. In worst case its complexity is $O(|\Omega|L^2)$, but the method appears to converge linearly in the number of parameter values in practice. An algorithm was proposed in [17] for segmentation problems with two regions, where a sequence of L total variation regularized problems could be solved, followed by L simple thresholding operations each step. The complexity is therefore effectively $O(|\Omega|L)$.

Contributions:

This work presents a jointly convex optimization framework for minimizing energy potentials of the form (4) over the regions and the parameters associated with each region (such as mean intensities). We also derive a convex relaxation of the piecewise constant Mumford-Shah model (7), where additionally the number of regions are unknown. The convex relaxation of (4) can be applied for problems with any number of regions, not just $n = 2$. Furthermore, the size of the convex relaxed problems grow at most linearly in the number of potential parameter values L, i.e. as $O(|\Omega|L)$.

The problems are first reformulated as minimization problems over binary functions defined in a space of one higher dimension than the image domain. Convex relaxations are then derived based on the reformulated problems. The method is not guaranteed to always produce an exact solution, but some conditions are identified for when this is possible. We begin by treating the piecewise constant Mumford Shah model in Section 2.2. Next, we present convex relaxations for the problems (4) and (8), where the number of regions are upper bounded. Fast algorithms are derived in Section 4.

2 Convex Relaxation Models

In this section the problems (4) and (7) are first reformulated as optimization problems in terms of a binary function in a space of one higher dimension than the image domain. Convex relaxations are then derived based on the reformulated problems. The new relaxations build on recently proposed convex relaxations for Potts model (1) which are briefly reviewed next.

2.1 Convex Relaxation for Potts Model

Several convex relaxations for Potts model (1) have recently been proposed [11, 22, 14, 2]. Any such convex relaxation can be used as building block for the new relaxations of the more complicated models (4) and (7) proposed in this work. However, we focus particularly on a simple relaxation for Potts model [11, 22, 2], which have demonstrated to work well for practical problems. Let $u_i(x)$ be the characteristic function of the region Ω_i, defined as

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

Then, the Potts model (1) can be written in terms of u_i as:

$$\min_{\{u_i\}_{i=1}^n \in B} \quad \sum_{i=1}^n \int_\Omega u_i(x) f_i(x)\, dx + \alpha \sum_{i=1}^n \int_\Omega |\nabla u_i|\, dx \tag{9}$$

subject to

$$\sum_{i=1}^n u_i(x) = 1, \quad \forall x \in \Omega \tag{10}$$

where B is the set

$$B = \{u \in BV(\Omega) : \; u(x) \in \{0,1\}, \; \forall x \in \Omega\} \tag{11}$$

and the total-variation of the characteristic function $u_i(x)$ encodes the length of the boundary of the region Ω_i.

A convex relaxation of (9), was proposed and studied in [11, 22, 19, 2] by instead minimizing over the convex set

$$u_i \in B' = \{u \in BV(\Omega) : \; u(x) \in [0,1], \; \forall x \in \Omega\} \tag{12}$$

for $i = 1, ..., n$. If the solution of the relaxed problem is binary at all $x \in \Omega$, it is also a global minimum of (1). Otherwise different schemes were proposed [11, 22, 2] to generate a binary solution $\tilde{u}$, which may either be a global minimum or close approximation to a global minimum of (1). The simplest such rounding scheme is just:

$$\tilde{u}(x) = e_\ell(x), \text{ where } \ell = \arg\max_i u_i(x) \tag{13}$$

2.2 Convex relaxation for Piecewise-Constant Mumford-Shah Model

In this section, we show that the piecewise constant Mumford-Shah model (7) can be expressed as a special case of (9). In its most general form, the Mumford-Shah model seeks an approximation image I and a set of curves Γ which minimizes

$$\inf_{\Gamma,I} E_\lambda(\Gamma,I) = \int_\Omega |I^0(x) - I(x)|^\beta dx + \lambda \int_{\Omega\setminus\Gamma} |\nabla I|^2 dx + \alpha \int_\Gamma ds. \tag{14}$$

Its piecewise constant variant can be regarded as the limit model as the penalty parameter λ goes to infinity i.e.

$$\inf_{\Gamma,I} E_\infty(\Gamma,I) \tag{15}$$

Due to infinite weight on the term $\int_{\Omega\setminus\Gamma} |\nabla I|^2$, (15) enforces solutions $I(x)$ that are constant everywhere except for the discontinuity set Γ, i.e. the function $I(x)$ is piecewise constant. The discontinuity set Γ therefore splits the domain Ω into a set of subdomains, say n in number: $\{\Omega_i\}_{i=1}^n$. The number n is unknown in advance, and is part of the optimization process. The piecewise constant Mumford-Shah model can therefore equivalently be formulated as (7), which is optimized over the regions Ω_i for $i = 1,...,n$, the mean values μ_i of $I(x)$ within each region Ω_i for $i = 1,...,n$ and the number of regions n.

Alternatively, (7) can be formulated in terms of the characteristic functions $u_i(x)$ as:

$$\min_n \min_{\{u_i\}_{i=1}^n \in B} \min_{\{\mu_i\}_{i=1}^n \in X} \quad E(u,\mu,n) = \sum_{i=1}^n \int_\Omega u_i(x) |I^0(x) - \mu_i|^\beta \, dx + \alpha \sum_{i=1}^n \int_\Omega |\nabla u_i| \, dx \tag{16}$$

subject to

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

Assume now that the set of feasible values μ_i is finite, i.e. $X = \{\ell_1,...,\ell_L\}$. For instance X may consist of the set of quantized gray values: $X = \{1,...,L\}$. For each element $\ell_i \in X$ define the corresponding characteristic function $u_i(x) \in B$. We will show the piecewise constant Mumford-Shah model (7) can be written as a minimization problem over such a set $\{u_i\}_{i=1}^L$. More specifically, we show that the following minimization problem is equivalent to the piecewise constant Mumford-Shah model (16) if the feasible intensity values are restricted to $X = \{\ell_1,...,\ell_L\}$.

$$\min_{\{u_i\}_{i=1}^L \in B} E^{\text{ext}}(\{u_i\}_{i=1}^L) = \sum_{i=1}^L \int_\Omega u_i |I^0(x) - \ell_i|^\beta \, dx + \alpha \sum_{i=1}^L \int_\Omega |\nabla u_i(x)| dx \tag{17}$$

subject to

$$\sum_{i=1}^L u_i(x) = 1, \quad u_i(x) \geq 0, \quad \forall x \in \Omega,\ i = 1, ..., L. \tag{18}$$

The above energy has the same form as (9).

Proposition 1. *Given an optimum u^* of* (17)*. Let n^* be the number of indices i for which $u_i^* \not\equiv 0$. Define the set of indices $\{i_j\}_{j=1}^{n^*} \subset \{1,...,L\}$ where $u_{i_j}^* \not\equiv 0$. Then $(\{\ell_{i_j}\}_{j=1}^{n^*}, \{u_{i_j}^*\}_{j=1}^{n^*}, n^*)$ is a global optimum of the piecewise constant Mumford-Shah model* (16) *with $X = \{\ell_1, ..., \ell_L\}$.*

The proof is given in the appendix.

In view of Prop. 1, a convex relaxation of the piecewise constant Mumford-Shah model can be defined as the minimization of (17) over $B' = \{u \in BV(\Omega) : u(x) \in [0,1]\ \forall x \in \Omega\}$. It has the same form as the convex relaxed Potts model, which has already been studied in [11, 22, 2, 19].

The piecewise constant Mumford-Shah model (16) will naturally result in a sparse solution, where the number of 'active' regions n is relatively low in comparison to L. The regularization parameter α controls both regularity of the region boundaries and the number of regions. If $\alpha = 0$, the solution is just $I = I^0$ and the pixels are not grouped in any way, instead each pixel is regarded as a distinct region. Therefore, the Mumford-Shah model may result in more regions than desired unless α is set sufficiently high.

2.3 Jointly Convex Relaxation over Regions and Region Parameters

In this section we propose a convex relaxation for image segmentation models where the number of regions are fixed, e.g. (4) or (8). In many applications, the number of regions is known in advance, but the region parameters are unknown. This is for instance the case for segmentation problems with two regions where one wishes to distinguish foreground and background.

We start by writing out (4) in terms of the characteristic functions u_i of each region Ω_i as follows

$$\min_{\{u_i\}_{i=1}^n \in B} \min_{\{\mu_i\}_{i=1}^n \in X} \quad \sum_{i=1}^n \int_\Omega u_i(x) |I^0(x) - \mu_i|^\beta \, dx + \alpha \sum_{i=1}^n \int_\Omega |\nabla u_i| \, dx. \tag{19}$$

s.t. $\sum_{i=1}^n u_i(x) = 1$ for all $x \in \Omega$. In order to optimize (4) over the set $\mu_i \in X = \{\ell_1, ..., \ell_L\}$, $i = 1, ..., n$, we start by proposing two equivalent alternative reformulations of (4):

Alternative 1: For each binary-valued function u_i, $i = 1, ..., L$, define a binary variable $v_i \in \{0,1\}$, with the interpretation $v_i = 1$ if $u_i(x) \neq 0$ for some $x \in \Omega$ and $v_i = 0$ else. Then (4) can be formulated as

$$\min_{u,v} \quad E^{\text{ext}}(\{u_i\}_{i=1}^L) = \sum_{i=1}^L \int_\Omega u_i(x)|I^0(x) - \ell_i|^\beta + \alpha |\nabla u_i| \, dx \tag{20}$$

subject to

$$\sum_{i=1}^L u_i(x) = 1 \,, \quad \forall x \in \Omega \,, \tag{21}$$

$$\sum_{i=1}^L v_i \leq n \,, \tag{22}$$

$$u_i(x) \leq v_i \,, \quad \forall x \in \Omega \,, \; i = 1, ..., L \tag{23}$$

$$u_i(x) \in \{0,1\} \,, \quad \forall x \in \Omega \,, \; i = 1, ..., L \tag{24}$$

$$v_i \in \{0,1\} \,, \quad i = 1, ..., L \tag{25}$$

Alternative 2: The problem can also be formulated without the artificial variable v. Observe that by definition, $\sup_{x\in\Omega} u_i(x) \leq v_i$, therefore the constraints (22) and (23) can be shortened by $\sum_{i=1}^L \sup_{x\in\Omega} u_i(x) \leq n$, which is also convex. Therefore the problem can equivalently be formulated as

$$\min_{u} \quad E^{\text{ext}}(\{u_i\}_{i=1}^L) = \sum_{i=1}^L \int_\Omega u_i(x)|I^0(x) - \ell_i|^\beta + \alpha |\nabla u_i| \, dx \tag{26}$$

subject to

$$\sum_{i=1}^L u_i(x) = 1 \,, \quad \forall x \in \Omega \,, \tag{27}$$

$$\sum_{i=1}^L \sup_{x\in\Omega} u_i(x) \leq n \,, \tag{28}$$

$$u_i(x) \in \{0,1\} \,, \quad \forall x \in \Omega \,, \; i = 1, ..., L \tag{29}$$

The constraint (28) forces the solution to satisfy $u_i \equiv 0$ for all but at most n indices $i \in \{1, ..., L\}$. The next result shows that an optimum of (19), or equivalently of (4), can be obtained by finding an optimal solution u^* to either of the two above problems, (20) or (26), through the following proposition.

Proposition 2. *Given an optimum u^* of* (20) *or* (26). *Let n^* be the number of indices i for which $u_i^* \not\equiv 0$. Define the set of indices $\{i_j\}_{j=1}^{n^*} \subset \{1, ..., L\}$ such that $u_{i_j}^* \not\equiv 0$. Then $\{\ell_{i_j}\}_{j=1}^{n^*}, \{u_{i_j}^*\}_{j=1}^{n^*}$ is a global optimum to* (19) *with $X = \{\ell_1, ..., \ell_L\}$.*

Clearly, $n^* \leq n$, otherwise the constraints (28) or (23) would be violated. The rest follows by an identical proof to that of Prop. 1.

Both formulations (20) and (26) are nonconvex due to the binary constraints on u and v. Convex relaxations can instead be derived by replacing the binary constraints (24), (25) by

$$u_i(x) \in [0,1], \ \ \forall x \in \Omega, \ \ i = 1,...,L \tag{30}$$
$$v_i \in [0,1], \ \ i = 1,...,L. \tag{31}$$

The convex relaxations resulting from the two alternatives (20) and (26) are equivalent as they share the same set of minimizers u. If the computed solution of the relaxed problem is binary for all $x \in \Omega$, it is also globally optimal to the original problem. If not, a close approximation can be obtained by the binarization scheme (13).

3 Some optimality results

In general, the convex relaxations are not guaranteed to produce an exact global minimum, but will provide close approximations. In this section we derive some conditions under which an exact solution can be obtained. First, we show that under L^1 data fidelity, the optimal gray values belong to the set of gray values which are already contained in the image. Second, we show that in case of two regions ($n = 2$) a thresholding scheme for producing exact solutions can applied under some conditions.

3.1 L^1 Data Fidelity

Consider the models (1) and (7) with f_i, $i = 1,\dots,n$, given by (3) and $\beta = 1$, i.e. the L^1 fidelity term. Assume further the input image $I^0(x)$ is quantized and takes values in the set $\{\ell_1,...,\ell_L\}$. The next result shows that there exists optimal parameters μ_i, $i = 1,\dots,n$, that also take values in the same set $\{\ell_1,...,\ell_L\}$. Hence it suffices to optimize μ over the set $X = \{\ell_1,...,\ell_L\}$. This result has previously been shown by the two-region problems ($n = 2$) in [8].

Proposition 3. *Given $I^0 : \Omega \mapsto \{\ell_1,...,\ell_L\}$, and consider the data term* (3) *with $\beta = 1$ and $X = \mathbb{R}$. There exists globally optimal $(\{\Omega_i\}_{i=1}^n, \{\mu_i^*\}_{i=1}^n, n)$ to the Mumford-Shah model (7) or $\{\Omega_i\}_{i=1}^n, \{\mu_i^*\}_{i=1}^n$ to* (4)*, where $\mu_i^* \in \{\ell_1,...,\ell_L\}$ for $i = 1,...,n$.*

Proof. The proof is by induction. When restricted to two regions, $n = 2$, the result was proved in [8], Theorem 1. Assume the result holds for $n = k$, then there exists a globally optimal solution $\{\Omega_i\}_{i=1}^{k+1}, \{\mu_i\}_{i=1}^{k+1}, k+1$ to the Mumford-Shah model (7), or $\{\Omega_i\}_{i=1}^{k+1}, \{\mu_i\}_{i=1}^{k+1}$ to (4), where $\{\mu_i\}_{i=1, i\neq j}^{k+1} \in \{1,...,L\}$. We will show the result also holds for $n = k+1$. Pick any $j \in \{1,...,k+1\}$, and consider the image domain $\Omega \backslash \Omega_j$. Clearly, $\{\Omega_i\}_{i=1, i\neq j}^{k+1}, \{\mu_i\}_{i=1, i\neq j}^{k+1}, k$ is globally optimal to the Mumford Shah model in the domain $\Omega\backslash\Omega_j$. It remains to show that also $\mu_j \in \{1,...,L\}$. Pick any $\ell \neq j \in \{1,...,L\}$. Then $\{\Omega_i\}_{i=1,i\neq\ell}^{k+1}, \{\mu_i\}_{i=1,i\neq\ell}^{k+1}, k$ is globally optimal to the Mumford Shah model in the domain $\Omega\backslash\Omega_\ell$. By the induction hypotheses it is possible that $\{\mu_i\}_{i=1,i\neq\ell}^{k+1} \in \{1,...,L\}$, which implies there exists optimal $\mu_j \in \{1,...,L\}$.

3.2 Exactness of Relaxation for $n = 2$

The relaxations are not in general exact, but will produce solutions that are optimal or nearly optimal. In case $n = 2$, exact solutions can be generated under some conditions. It suffices that for two indices k, j, the boundary $u_k(x) = 1$ and $u_j(y) = 1$ is attained for some $x, y \in \Omega$.

Proposition 4. *Let u^* be a solution of* (26)*, or alternatively u^*, v^* a solution of* (20) *with $n = 2$, where the binary constraints B are replaced by the convex constraint B'. Assume the variable v is binary, or equivalently, assume there exists $k, j \in \{1,...,L\}$ such that $u_k(x) = 1$ for some $x \in \Omega$ and $u_k(y) = 1$ for some $y \in \Omega$ For any threshold level $t \in (0,1)$ define the function $\tilde{u}$ such that*

$$\tilde{u}_k(x) := \begin{cases} 1\,, & \text{if } u_i^*(x) \geq t \\ 0\,, & \text{if } u_i^*(x) < t \end{cases}, \quad \tilde{u}_j(x) := \begin{cases} 1\,, & \text{if } u_i^*(x) > 1-t \\ 0\,, & \text{if } u_i^*(x) \leq 1-t \end{cases}.$$

and $\tilde{u}_i = u_i$ for all $i \neq k, j \in \{1,...,L\}$. Then $(\tilde{u}, v^*)$ is a binary global optimum of (20) *subject to* (21)-(23) *and the binary constraints* (24) *and* (25)*.*

Proof. Since $u_k(x) = 1$ for some $x \in \Omega$ and $u_k(y) = 1$ for some $y \in \Omega$, it follows by constraint (28) that $u_i(x) = 0$ for all $i \neq k, j \in \{1,...,L\}$. Define $\phi = u_k$, then since $u_k + u_j = 1$, $\phi = 1 - u_j$. Define

$$\tilde{\phi}(x) := \begin{cases} 1\,, & \phi(x) \geq t \\ 0\,, & \phi(x) < t \end{cases}.$$

and observe that $\tilde{\phi} = \tilde{u}_k$ and $\tilde{\phi} = 1 - \tilde{u}_j$. Then

$$\begin{aligned} E^{\text{ext}}(u) &= \int_\Omega u_k(x)|I^0(x) - k|^\beta + u_j(x)|I^0(x) - j|^\beta + \alpha \int_\Omega |\nabla u_k| + |\nabla u_j|\, dx \\ &= \int_\Omega \phi(x)|I^0(x) - k|^\beta + (1 - \phi(x))|I^0(x) - j|^\beta + 2\alpha \int_\Omega |\nabla \phi|\, dx \\ &= \int_\Omega \tilde{\phi}(x)|I^0(x) - k|^\beta + (1 - \tilde{\phi}(x))|I^0(x) - j|^\beta + 2\alpha \int_\Omega |\nabla \tilde{\phi}|\, dx \\ &= \int_\Omega \tilde{u}_k(x)|I^0(x) - k|^\beta + \tilde{u}_j(x)|I^0(x) - j|^\beta + \alpha \int_\Omega |\nabla \tilde{u}_k| + |\nabla \tilde{u}_j|\, dx = E^{\text{ext}}(\tilde{u}). \end{aligned}$$

The third equality follows by the thresholding theorem of [7] for relaxed binary segmentation problems.

4 Algorithms

The convex relaxation for the piecewise constant Mumford-Shah model (17) has the form of the convex relaxed Potts model [10, 22], and can be optimized by established algorithms. In [19] a very efficient algorithm was proposed based on the dual formulation, which can also be parallelized over each characteristic function. This algorithm is therefore well suited for optimizing (17), which usually contains a large number of characteristic functions.

The convex relaxation of (4) is a little more complicated due the the extra constraints. As stated in Section 2.3, there are two equivalent formulations of the relaxation. We will build up an algorithm based on alternative 2. We assume the optimal number of regions n is attained (i.e. equality in (28)). If the optimal number of regions is less than n, exactly the same solution would be produced by the convex relaxation of the piecewise constant Mumford-Shah model, which is simpler to optimize and could be checked by a separate calculation. Let γ be a Lagrange multiplier for the constraint

$$\sum_{i=1}^{L} \sup_{x\in\Omega} u_i(x) - n = 0. \tag{32}$$

The problem can then be stated as the saddle point problem

$$\max_{\gamma} \min_{u} \mathscr{L}(u,\gamma) = \sum_{i=1}^{L} \int_{\Omega} u_i(x)|I^0(x) - \ell_i|^{\beta} + \alpha\,|\nabla u_i|\,dx + \gamma(\sum_{i=1}^{L} \max_{x\in\Omega} u_i(x) - n) \tag{33}$$

$$\text{s.t.} \quad \sum_{i=1}^{L} u_i(x) = 1, \quad u_i(x) \geq 0\ \forall x\in\Omega,\ i=1,...,L, \quad \gamma \geq 0$$

In order to optimize (33), the Lagrangian method can be applied as follows: for $k = 1,...$ until convergence

1. $u^{k+1} = \arg\min_u \mathscr{L}(u,\gamma^k), \quad \text{s.t. } \sum_{i=1}^{L} u_i(x) = 1, \quad u_i(x) \geq 0\ \forall x\in\Omega,\ i=1,...,L$
2. $\gamma^{k+1} = \max(0, \gamma^k + c(\sum_{i=1}^{L} \max_{x\in\Omega} u_i^{k+1}(x) - n))$.

Observe that subproblem 1. has the same form as the label cost prior problem studied in [20, 18]. A fast algorithm for solving such problems was proposed in [18]. In particular, it was shown 1. could be written as the primal-dual problem

$$\min_{u \in B', \sum_{i=1}^{L} u_i(x)=1, \forall x \in \Omega} \mathscr{L}(u, \gamma^k) = \min_{u} \max_{p_s, p_t, q, r} \int_\Omega p_s \, dx + \sum_{i=1}^{n} \int_\Omega u_i (\operatorname{div} q_i - p_s + p_i - r_i) \tag{34}$$

$$\text{s.t.} \quad p_i(x) \le |I^0(x) - \ell_i|^\beta \, , \; |q_i(x)| \le \alpha \, , \int_\Omega |r_i(x)| \, dx \le \gamma^k \, ; \; i = 1 \ldots n \, .$$

where u_i works as a Lagrange multiplier. The above energy functional can be optimized separately for p_s, p_i, q_i and r_i in closed form. Therefore the augmented Lagrangian method could be applied to efficiently solve the overall problem. In practice, only a few iterations are necessary before γ is updated.

5 Numerical Experiments

In this section we demonstrate numerically the new convex relaxation for optimizing the energy in the image segmentation model (4) jointly over the regions and regions parameters, and the new convex relaxation of the piecewise constant Mumford-Shah model (7). In Figures 1, 2 and 4, we have used 100 quantization levels for the unknown parameters, i.e. $X = \{0.01, 0.02, ..., 1.00\}$ and in Figure 3, 255 levels have been used. In order to visualize results, we depict the function $u(x) = \ell_i$ if $x \in \Omega_i$, $i = 1, ..., n$.

Observe that the piecewise constant Mumford-Shah model may result in more regions than desired, as shown in the last subfigures. This is especially visible in Figure 1 and 2, whereas it leads to more reasonable results in Figure 3 and 4. By instead minimizing (4), with the number of regions fixed to 2, in terms of the regions and parameters μ_1 and μ_2, one is able to separate foreground and background in Figure 1 and Figure 2. Observe that the piecewise constant Mumford-Shah model leads to lower energy, since it is optimized over a larger feasible set.

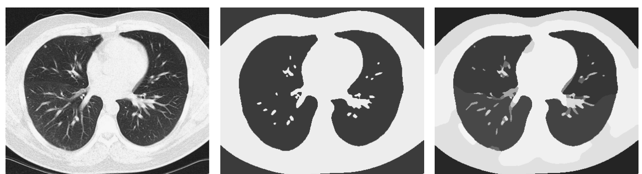
The convex relaxations generate close approximations to a global minimum. To verify this, we have used the estimated parameters μ_1 and μ_2 from the convex relaxation as initialization of the alternating minimization algorithm (5) - (6). In all cases, the converged values of μ_1 and μ_2, after rounding to the nearest element in X, did not change. This indicates strongly that the globally optimal values of μ_1 and μ_2 within X, had been obtained by the convex relaxation method. In our experience, the alternating algorithm (5) - (6) is rather robust to initialization and converges to the same solution for many initializations of μ_1 and μ_2. However, an independent work [4] presented examples where the alternating algorithm gets stuck in poor local minima for exactly the input images in Figure 1 and 2.

6 Conclusions and future work

We end with some discussions on future work and conclusions.

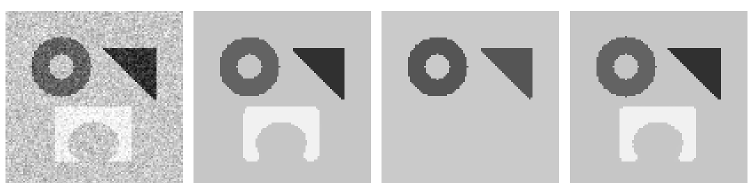
(a) (b) (c) (d)

Fig. 1 (a) Input image. (b)-(c) Convex relaxation of (4) with $n = 2$ and $\alpha = 0.15$: (b) $\beta = 2$, estimated parameters $\mu_1 = 0.09, \mu_2 = 0.59$, energy $= 1.25 * 10^3$; (c) $\beta = 1$ estimated parameters $\mu_1 = 0.06, \mu_2 = 0.62$, energy $= 5.39 * 10^3$. (d) Convex relaxation of piecewise constant Mumford-Shah model (7) with $\beta = 1$ and $\alpha = 0.15$, energy $= 3.31 * 10^3$.



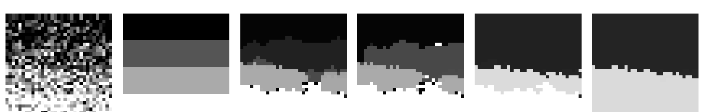
(a) (b) (c)

Fig. 2 (a)Input image. (b) Convex relaxation of (4) with $n = 2$, $\beta = 1$ and $\alpha = 0.15$: estimated parameters $\mu_1 = 0.23, \mu_2 = 0.93$, energy $9.23 * 10^3$. (d) Convex relaxation of piecewise constant Mumford-Shah model (7) with $\beta = 1$, energy $= 7.13 * 10^3$.



(a) (b) (c) (d)

Fig. 3 (a) Input. (b)-(c) Convex relaxation of (4): (b) $n = 4$, (c) $n = 2$. (d) Convex relaxation of piecewise constant Mumford-Shah model (7).



(a) (b) (c) (d) (e) (f)

Fig. 4 (a) Input. (b) Ground truth. (c) Convex relaxation of Mumford-Shah functional: energy 212.02. (d)-(f) Convex relaxation of (4): (d) $n = 4$, (e) $n = 3$, (f) $n = 2$ energy $= 219.78$. In all experiments $\beta = 1$.

6.0.1 Extension to continuous label values

The convex optimization framework for (4) and requires that the set of feasible parameter values is quantized. The relaxations can also be extended to optimization problems where the set of feasible parameter values is continuous, i.e. $X = \mathbb{R}$. Let $\ell_{\min}$ be the smallest and $\ell_{\max}$ be the largest value of μ_i. Define the one-higher dimensional binary variable $u(x,\ell)$ for each $(x,\ell) \in \Omega \times [\ell_{\min}, \ell_{max}]$, i.e. $u : \Omega \times [\ell_{\min}, \ell_{max}] \mapsto \{0,1\}$.

As a continuous generalization of (17), we argue the piecewise constant Mumford-Shah model (7) can be formulated in terms of u as

$$\min_u \quad \int_{\ell_{\min}}^{\ell_{\max}} \int_{\Omega} u(x,\ell)|I^0(x) - \ell|^{\beta} + \alpha |\nabla_x u(x,\ell)| \, dx d\ell \tag{35}$$

subject to

$$\int_{\ell_{\min}}^{\ell_{\max}} u(x,\ell) * \delta(\ell) \, d\ell = 1, \quad \forall x \in \Omega \tag{36}$$

$$u(x,\ell) \in \{0,1\}, \quad \forall (x,\ell) \in \Omega \times [\ell_{\min}, \ell_{\max}]. \tag{37}$$

where $\delta(\ell)$ is the delta distribution and the convolution is defined as $u(x,\ell) * \delta(\ell) = \int_{-\infty}^{\infty} u(x,\ell)\delta(\ell - s) ds$.

Let $u^*(x,\ell)$ be an optimum of (35). We conjecture that $w^* = \int_{-\infty}^{\infty} \ell u^*(.,\ell) * \delta(\ell) \, d\ell$ is a piecewise constant function that is a global minimizer of the piecewise constant Mumford-Shah model (15). We believe the proof can be constructed as a direct continuous generalization of the proof of Prop. 1, but will be more involved due to measure theoretic aspects. It would be interesting to investigate how this result relates to a recently proposed convex relaxation of the piecewise smooth Mumford-Shah model [15].

6.0.2 Extension to Vector-Valued Parameters

The results discussed in Sec. 2.3 can easily be extended to more general problems of the form (8), where $\xi = (\xi_1, ..., \xi_N)$ denote the vector-valued parameter associated with each region. Let $X = \{\ell_1^1, ..., \ell_L^1\} \times ... \times \{\ell_1^N, ..., \ell_L^N\}$ denote the finite set of all feasible ξ. For each $i_1, i_2, ..., i_N \in \{1, ..., L\}$ define the function $u_{i_1,...,i_N} : \Omega \mapsto \{0,1\}$ and variable $v_{i_1,...,i_N} \in \{0,1\}$. Then the model (8) can be written

$$\min_u \quad \sum_{i_1=1}^{L} ... \sum_{i_N=1}^{L} \int_{\Omega} u_{i_1,...,i_N}(x) f(\xi_{i_1}^1, ..., \xi_{i_N}^N, x) + \alpha |\nabla u_{i_1,...,i_N}| \tag{38}$$

subject to

$$\sum_{i_1=1}^{L} \cdots \sum_{i_N=1}^{L} u_{i_1,\ldots,i_N}(x) = 1\,, \quad \forall x \in \Omega \tag{39}$$

$$\sum_{i_1=1}^{L} \cdots \sum_{i_N=1}^{L} \max_{x \in \Omega} u_{i_1,\ldots,i_N}(x) \leq n, \tag{40}$$

$$u_{i_1,\ldots,i_N}(x) \in \{0,1\}\,, \quad \forall x \in \Omega,\; i_1,\ldots,i_N \in X\,. \tag{41}$$

The equivalence between (8) and (38) follows by a straight forward generalization of Prop. 2.

6.1 Conclusions

Image segmentation problems can successfully be modeled as the minimization of an energy potential with respect to regions and parameters associated with each region. In this work, we have reformulated such problems as the optimization of binary functions in a space of one higher dimension than the image domain. Convex relaxations and optimization algorithms have been proposed which does not depend on initializations and produce close approximations to global minima. In contrast to previous work, the complexity of our algorithm grows at most linearly with the number of potential parameter values, and can be applied for segmentation problems with any number of regions.

7 Proofs

Proof of Prop 1

Proof. Let $(\{\tilde{u}_j\}_{j=1}^{\tilde{n}}, \{\ell_{\tilde{i}_j}\}_{j=1}^{\tilde{n}}, \tilde{n})$ be any other solution of (16). Define the vector function

$$\bar{u}_j = 0, \text{ for } j \in \{1,\ldots,L\} \backslash \{\tilde{i}_1,\ldots,\tilde{i}_{\tilde{n}}\}$$
$$\bar{u}_{\tilde{i}_j} = \tilde{u}_j \text{ for } j = 1,\ldots,\tilde{n}.$$

Then $\bar{u}$ belongs to the feasible set (18) of the problem (17).

$$E^{\text{ext}}(\bar{u}) = \sum_{i=1}^{L} \int_\Omega \bar{u}_i |I^0(x) - \ell_i|^2\, dx + \sum_{i=1}^{L} \alpha \int_\Omega |\nabla \bar{u}_i|\, dx = \sum_{j=1}^{\tilde{n}} \int_\Omega \bar{u}_{\tilde{i}_j} |I^0(x) - \ell_i|^2 dx + \sum_{j=1}^{\tilde{n}} \alpha \int_\Omega |\nabla \bar{u}_{\tilde{i}_j}| dx$$

$$= \sum_{i=1}^{\tilde{n}} \int_\Omega \tilde{u}_i |I^0(x) - \ell_{\tilde{i}_j}|^2\, dx + \sum_{i=1}^{\tilde{n}} \alpha \int_\Omega |\nabla \tilde{u}_i|\, dx = E(\{\tilde{u}_j\}_{j=1}^{\tilde{n}}, \{\ell_{\tilde{i}_j}\}_{j=1}^{\tilde{n}}, \tilde{n}).$$

But since u^* is a global minimizer of E^{ext}

$$E^{\text{ext}}(u^*) \leq E^{\text{ext}}(\bar{u}) = E(\{\tilde{u}_j\}_{j=1}^{\tilde{n}}, \{\ell_{\tilde{i}_j}\}_{j=1}^{\tilde{n}}, \tilde{n}), \tag{42}$$

and since

$$E^{\text{ext}}(u^*) = \sum_{i=1}^{L} \int_\Omega u_i^* |I^0(x) - \ell_i|^2 \, dx + \sum_{i=1}^{L} \alpha \int_\Omega |\nabla u_i^*| \, dx$$

$$= \sum_{j=1}^{n} \int_\Omega u_{i_j}^* |I^0(x) - \ell_{i_j}|^2 dx + \sum_{j=1}^{n} \alpha \int_\Omega |\nabla u_{i_j}^*| dx = E(\{u_{i_j}^*\}_{j=1}^{n}, \{i_j\}_{j=1}^{n}, n). \tag{43}$$

Combining (42) and (43) it follows that

$$E(\{u_{i_j}^*\}_{j=1}^{n}, \{i_j\}_{j=1}^{n}, n) \leq E(\{\tilde{u}_j\}_{j=1}^{\tilde{n}}, \{\ell_{\tilde{i}_j}\}_{j=1}^{\tilde{n}}, \tilde{n}).$$

Hence $\{u_{i_j}^*\}_{j=1}^{n}, \{i_j\}_{j=1}^{n}, n$ must be a solution to (16).

References

1. Egil Bae and Xue-Cheng Tai. Efficient global minimization for the multiphase Chan-Vese model of image segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 28–41, 2009.
2. Egil Bae, Jing Yuan, and Xue-Cheng Tai. Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision*, 92(1):112–129, 2011.
3. Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:1222 – 1239, 2001.
4. Ethan S. Brown, Tony F. Chan, and Xavier Bresson. Completely convex formulation of the chan-vese image segmentation model. *International Journal of Computer Vision*, 2011. doi: 10.1007/s11263-011-0499-y.
5. Ethan S. Brown, Tony F. Chan, and Xavier Bresson. A convex relaxation method for a class of vector-valued minimization problems with applications to mumford-shah segmentation. *UCLA, Applied Mathematics, CAM-report-10-43*, July, 2010.
6. T. Chan and L.A. Vese. *Active contours without edges*. IEEE Image Proc., 10, pp. 266-277, 2001.
7. Tony F. Chan, Selim Esedoḡlu, and Mila Nikolova. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.*, 66(5):1632–1648 (electronic), 2006.
8. J. Darbon. A note on the discrete binary mumford-shah model. *Proceedings of Computer Vision/Computer Graphics Collaboration Techniques, (MIRAGE 2007), LNCS Series vol. 44182*, pages 283–294, March 2007.
9. Andrew Delong, Anton Osokin, Hossam Isack, and Yuri Boykov. Fast approximate energy minimization with label costs. In *CVPR*, 2010.
10. J. Lellmann, D. Breitenreicher, and C. Schnörr. Fast and exact primal-dual iterations for variational problems in computer vision. *European Conference on Computer Vision (ECCV)*, 2010.
11. J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. In X.-C. Tai, K. Mórken, M. Lysaker, and K.-A. Lie, editors, *Scale Space and Variational Methods in Computer Vision (SSVM 2009)*, volume 5567 of *LNCS*, pages 150–162. Springer, 2009.

12. Victor Lempitsky, Andrew Blake, and Carsten Rother. Image segmentation by branch-and-mincut. In *Proceedings of the 10th European Conference on Computer Vision: Part IV*, pages 15–29, Berlin, Heidelberg, 2008. Springer-Verlag.
13. D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math, 42*, 42:577–685, 1989.
14. T. Pock, A. Chambolle, H. Bischof, and D. Cremers. A convex relaxation approach for computing minimal partitions. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, Florida, 2009.
15. T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth mumford-shah functional. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.
16. Renfrey B. Potts. Some generalized order-disorder transformations. *In Proceedings of the Cambridge Philosophical Society, Vol. 48*, pages 106–109, 1952.
17. Petter Strandmark, Fredrik Kahl, and Niels Chr Overgaard. Optimizing parametric total variation models. In *ICCV*, pages 26–33, 2009.
18. J. Yuan, E. Bae, Y. Boykov, and X.C. Tai. A continuous max-flow approach to minimal partitions with label cost prior. In *Scale Space and Variational Methods in Computer Vision*, 2011.
19. Jing Yuan, Egil Bae, Xue-Cheng Tai, and Yuri Boykov. A continuous max-flow approach to potts model. In *ECCV*, 2010.
20. Jing Yuan and Yuri Boykov. Tv-based image segmentation with label cost prior. In *BMVC*, 2010.
21. Jing Yuan, Juan Shi, and Xue-Cheng Tai. A convex and exact approach to discrete constrained tv-l1 image approximation. Technical report CAM-10-51, UCLA, CAM, UCLA, 2010.
22. 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.
23. Song Chun Zhu and Alan Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.