Mumford-Shah type variational models and fast algorithms for image segmentation

Gu, Ying

2013

Gu, Y. (2013). Mumford-Shah type variational models and fast algorithms for image segmentation. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/51190>

<https://doi.org/10.32657/10356/51190>

MUMFORD-SHAH TYPE VARIATIONAL MODELS AND FAST ALGORITHMS FOR IMAGE SEGMENTATION

GU YING

School of Physical and Mathematical Sciences

A thesis submitted to the Nanyang Technological University in
fulfilment of the requirement for the degree of
Doctor of Philosophy

2013

Acknowledgement

First and foremost, I would like to thank my supervisor Prof. Li-Lian Wang for the guidance and providing me the privilege of working in this interesting research area. Without his invaluable supervision, encouragement, and wisdom over these years, this thesis could not have reached its present form. At the same time, I am also grateful to Prof. Xue-Cheng Tai for providing me with valuable advice and access to the related resources on my thesis.

I am grateful to Nanyang Technological University for awarding me the Research Scholarship. I am also benefited from the young, energetic and productive environment of the new School of Physical and Mathematical Sciences, where I could grow and develop. In particular, I appreciate the management and leadership of Prof. San Ling, Prof. Yeow Meng Chee, Prof. Bernhard Schmidt and Prof. Chaoping Xing. I am appreciative of the help from the IT and Admin. staff.

I would like to express my appreciation to Prof. Desheng Wang and Prof. Viet Ha Hoang for the stimulating lectures and seminars. I am also indebted to other members in our group including Jie Chen, Yuping Duan, Jing Zhang, Michael Daniel Villena Samson, Juan Shi, Min Wan, Yu Wang and Xiaodan Zhao, for their friendship and help.

Last but not least, I wish to thank my parents and my brother for their love and support.

Contents

Acknowledgement	3
List of Figures	10
List of Tables	11
Abstract	13
Notation	17
1 Introduction	1
1.1 Background and Motivations	2
1.1.1 Mumford-Shah model	2
1.1.2 Piecewise constant Mumford-Shah model	5
1.1.3 Potts model	6
1.1.4 Level-set methods	7
1.2 Main Contributions of the Dissertation	14
2 Efficient Dual Algorithms for TV-Allen-Cahn Model	17
2.1 Splitting and Penalty Technique	20
2.2 Dual Algorithm based on Chambolle [34]	24
2.3 Edge-Featured Penalty and FFT	26
2.4 Numerical Results	29

2.5	Applications to Image Denoising	34
3	Global Minimization for Multiphase Labeling and Segmentation	
	Problems	37
3.1	Analysis of the Primal-Dual Model and the Algorithm	40
3.1.1	Two-phase case	40
3.1.2	Multiphase case	44
3.1.3	Description of the algorithm	47
3.2	A Comparison Study and Numerical Results	51
3.2.1	A comparison study	51
3.2.2	Numerical tests	55
4	Robust Edge Detection Using A Modified Mumford-Shah Model	73
4.1	The Modified Mumford-Shah Model	75
4.2	Minimization Algorithm	77
4.2.1	FFT for solving u	79
4.2.2	Proximity algorithm for solving ψ	79
4.2.3	Split Bregman method for solving ψ	84
4.3	Numerical Experiments	86
4.3.1	Algorithm for the Ambrosio and Tortorelli model	86
4.3.2	Selection of the parameters α and β	89
4.3.3	More tests and extension to color images	89
5	A Decomposition Model for Piecewise Smooth Image Segmentation	95
5.1	The Decomposition Model	97
5.2	Relaxation and Well-posedness	98
5.3	Minimization Algorithm	101
5.3.1	Minimization with respect to b	102
5.3.2	Smoothed Chambolle-Dual algorithm for solving $\mathbf{v}$	102
5.3.3	Proximity algorithm for solving $\mathbf{v}$	105

5.4	Numerical Results	107
5.4.1	One-dimensional tests	108
5.4.2	The Vese-Chan method [135]	108
5.4.3	More two-dimensional tests	112
6	Conclusion and Future Works	115
6.1	Conclusion	115
6.2	Future Works	117
	Bibliography	119
	List of Publications	133

List of Figures

1.1	Labeling of four subdomains by two signed distance functions.	9
1.2	Signed distance function and level-set function	11
2.1	Profiles of the “double-well” potentials	23
2.2	Profiles of $ \nabla I $, $ \nabla \psi $ and $ \mathbf{p} $ obtained by Algorithm 1	27
2.3	Comparison of the convergence rate for three different time step sizes .	31
2.4	Comparison results with two different choices of initial values	32
2.5	Segmentation by Algorithms 1 and 2	33
2.6	Segmentation by Algorithms 1 and 2 with the same parameters	33
2.7	Segmentation by Algorithms 2	34
2.8	Results obtained by different algorithms	36
3.1	Comparison of five algorithms	57
3.2	Comparison of Algorithms 1-2 and GSD method for two-phase case . .	60
3.3	Comparison of decay of numerical energy for Set 1-3	61
3.4	Comparison of Algorithms 2 and GSD algorithm for four-phase case .	61
3.5	Comparison of decay of numerical energy for test Set 4 and Set 5 . .	62
3.6	Comparison of different algorithms	63
3.7	Completion of edges	63

3.8	Results obtained by Algorithms 2 with different β	65
3.9	Decay of numerical energy for different β	66
3.10	Results obtained by Algorithms 2	67
3.11	Results obtained by Algorithms 2	68
3.12	Comparison of Algorithm 3 and GSD-Seg	70
3.13	More tests on Algorithm 3	71
4.1	Examples of different curves	76
4.2	Comparison of four algorithms for clean images	89
4.3	Comparison of four algorithms for noisy images	90
4.4	Results obtained by FFT-P algorithm with different α and β	91
4.5	Results obtained by FFT-P algorithm and FFT-SP algorithm	92
4.6	Results obtained by FFT-SP algorithm	94
5.1	Results obtained by FFT-SCD algorithm and FFT-PS algorithm	109
5.2	Comparison of three different algorithms	111
5.3	Results obtained by FFT-SCD algorithm and FFT-PS algorithm	113
5.4	Results obtained by FFT-SCD algorithm and FFT-PS algorithm	114
6.1	Results obtained by the integral smoothing approach	118

List of Tables

3.1	Comparison of five algorithms.	56
3.2	Comparison of Algorithm 2 and GSD.	62
3.3	Comparison of Algorithm 3 and GSD-Seg.	70
5.1	Comparison of three different algorithms.	112

Abstract

Image segmentation, which aims to extract interesting objects from a given image, is one fundamental task in image processing and computer vision. Among several popular variational models for this purpose, we focus on the seminal Mumford and Shah (MS) model in this thesis.

Build upon the MS model, we present various modifications that adapt to different segmentation tasks, and develop efficient minimization algorithms for the proposed modified models. Below, we guide you through the main contents and contributions of this dissertation.

- (i) We develop efficient dual-algorithm for two-phase image segmentation using TV-Allen-Cahn model (see Chapter 2). The main purpose is to overcome the difficulties resulted from the non-differentiability of the total variation (TV) term, and effectively handle the constraints of the binary level-set function. The use of a splitting-penalty method results in TV-Allen-Cahn type models associated with a very attractive “double-well” potential, which allows for the implementation of the Chambolle’s dual algorithm. Moreover, we present a new dual algorithm based on an edge-featured penalty of the dual variable, which only requires to solve a vectorial Allen-Cahn type equation with linear $\nabla(\text{div})$ -diffusion rather than fully nonlinear diffusion in the Chambolle’s approach. Consequently, efficient numerical algorithms such as time-splitting method and Fast Fourier Transform (FFT) can be implemented. Various numerical tests show that two dual algorithms are much faster and stabler than the primal gradient descent approach, and the new

dual algorithm is at least as efficient as the Chambolle's algorithm and is relatively accurate. We demonstrate that the new method also provides a viable alternative for image restoration.

- (ii) We propose a direct global minimization method for multiclass labeling and multiphase image segmentation (see Chapter 3). Different from the existing methods, we work directly with the binary setting without using convex relaxation, which is thereby termed as a direct approach. We provide the sufficient and necessary conditions to guarantee a global optimum, and present efficient algorithms based on a reduction of the intermediate unknowns from the augmented Lagrangian formulation. As a result, the underlying algorithms involve significantly fewer parameters and unknowns than the naive use of augmented Lagrangian-based methods, so they are fast and easy to implement. Furthermore, they can produce a global optimum under mild conditions.
- (iii) We present a robust edge detection method by using the modified Mumford-Shah model (see Chapter 4). This model immerses an edge into a narrow band surrounding it, and measures the edge length by TV of the associated binary level-set function. The situation that the edge set could be of measure zero in the limiting process presents significant challenges for minimization. We propose several ideas to surmount the obstacles, which include the convex relaxation technique, splitting-penalty method, the proximity algorithm and split Bregman method. The approach can also be extended to color image processing, where limited results are available.
- (iv) We propose a new decomposition model for multiphase piecewise smooth image segmentation (see Chapter 5). By decomposing an image into an interpolation of piecewise constant part, smooth component and noise part, we incorporate these components into the Mumford-Shah model. This only requires to solve the smooth component on the whole domain. With the

entropy penalization, we construct efficient algorithms which include the Fourier method, smoothed Chambolle-dual algorithm and proximity algorithm. Numerical results are provided to show the advantages over the existing method.

Notation

$\mathbb{R}$	the set of all real numbers
$\mathbb{Z}^+$	the set of all positive integers
$\mathbb{N}$	the set of all natural numbers

Let $d \in \mathbb{Z}^+$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, $\boldsymbol{x} = (x_1, \dots, x_d)$, $\boldsymbol{p} = (p_1, \dots, p_d) \in \mathbb{R}^d$

$ \boldsymbol{\alpha} $	$ \boldsymbol{\alpha} = \sum_{i=1}^d \alpha_i$
∂_i	partial derivative with respect to x_i , $\partial_i = \partial/\partial x_i$
D^α	$D^\alpha f = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d} f$
∇	the gradient operator, $\nabla f = (\partial_1 f, \dots, \partial_d f)$
div	the divergence operator, $\text{div} \boldsymbol{p} = \partial_1 p_1 + \dots + \partial_d p_d$

Ω	image domain
Γ	the boundary of Ω , $\Gamma = \partial\Omega$
$\boldsymbol{n}$	the unit outer normal vector to $\partial\Omega$
$C^\infty(\Omega)$	the space of infinitely differentiable functions
$C^k(\Omega)$	the space of differentiable functions whose partial derivatives up to k th order are continuous
$C_c^k(\Omega)$	the space of functions in $C^k(\Omega)$ with compact support
$BV(\Omega)$	the space of functions of bounded variations on Ω

$L^p(\Omega)$	the L^p -space (see Adams [1]), $1 \leq p \leq \infty$, with the norm $\ \cdot\ _{L^p(\Omega)}$
$W^{m,p}(\Omega)$	the Sobolev space as in [1], with the norm $\ \cdot\ _{W^{m,p}(\Omega)}$
I	a given image
u	smooth approximation of I
$\lambda, \alpha, \beta, \mu, \nu$	weighted and tuning parameters
ϕ or $\{\phi_i\}$	signed distance function
ψ or $\{\psi_i\}$	level-set function
χ	$\chi = (\chi_1, \chi_2, \dots, \chi_m)$, the characteristic functions

Introduction

Image segmentation aims to partition a given image into a number of constituent regions so as to extract objects of specific, distinctive features from image background. Indeed, image segmentation is a fundamental task in image processing and computer vision, and is found in ubiquitous applications (see, e.g., [2, 99]).

Variational approaches have been very influential and powerful methods for image segmentation. In a variational framework, image segmentation is achieved by solving a certain minimization problem involving a cost functional (or energy functional) that mathematically incorporates various necessary a priori or a posteriori information of the given and targeted images. The popular variational models for image segmentation include the region-based Mumford-Shah model [95], the edge-based snakes of Kass, Witkin and Terzopoulos [62], and their important variants and modifications, e.g., the Chan-Vese model of active contour without edges [42] and the geodesic active contour of Caselles, Kimmel and Sapiro [31].

The multiclass labeling or classification problem, which is intimately related to (multiphase) image segmentation, intends to classify a pixel into a specific class according to some optimization criterion. The problem is known to be NP-hard if it is formulated and solved in a discrete manner [64, 107, 104, 24, 14]. This drives the minimization of the (continuous) Potts model [108], which shares a

very close cost functional with the piecewise constant Mumford-Shah model.

It is known that most of the variational models for image segmentation and classification are nonconvex in either the involved energy functional or the constraint optimization space, so the practitioners and researchers are oftentimes plagued with unpleasant local minima solutions or unsatisfactory algorithms that are sensitive to initial inputs. Hence, the search for fast global minimization algorithms for such problems has been a longstanding issue among this community.

This thesis tackles the image segmentation and multiclass labeling problems based on the Mumford-Shah model, and contributes to the development of new modified Mumford-Shah models (termed as “Mumford-Shah-type models”), and the construction of efficient algorithms for the proposed models. Below, we provide a background study and highlight our motivations and main contributions.

1.1 Background and Motivations

In this section, we outline some variational models and relevant mathematical preliminaries essential to this dissertation, and more importantly, we emphasize the major motivations of this study.

1.1.1 Mumford-Shah model

We start with introducing the seminal model proposed by Mumford and Shah in 1989 [95]. Let Ω be an open bounded domain in $\mathbb{R}^2$, and $I : \Omega \rightarrow \mathbb{R}$ be a given image to be segmented. We search for an optimal piecewise smooth approximation u of I , and an optimal edge set $\Gamma(\subset \Omega)$ that separates the piecewise smooth regions. This can be formulated as the minimization problem:

$$\min_{u, \Gamma} \left\{ E_{MS}(u, \Gamma) := \mu \int_{\Omega \setminus \Gamma} |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + |\Gamma| \right\}, \quad (1.1)$$

where $|\Gamma|$ is the length of Γ , and $\mu, \nu > 0$ are parameters to weight the terms in the functional. The term $\int_{\Omega} |u - I|^2 d\mathbf{x}$, known as the data fidelity term, forces

u to be close to the given image I , while the term $\int_{\Omega \setminus \Gamma} |\nabla u|^2 d\mathbf{x}$ penalizes strong variations of u , thus ensures that u is a smooth approximation of I . Through penalizing the length of discontinuities, the model also favors to find a “tight” enclosure of the smooth regions. It is seen that this model involves two unknowns with different nature, and this presents a great challenge for mathematical and numerical analysts. We refer to [47] and the references therein for a review of many insightful analysis and various attempts of approximating, relaxing or simplifying the Mumford-Shah functional.

At this point, an important issue is how to represent the edge set and how to relate it to the unknown u . In the context of image segmentation, we consider the edge set consisting of rectifiable closed curves [94]. Accordingly, the edge set Γ partitions Ω into m disjoint Ω_i , i.e.,

$$\Omega = \bigcup_{i=1}^m \Omega_i, \quad \Omega_i \cap \Omega_j = \emptyset \text{ for } i \neq j, \quad (1.2)$$

and I is approximated by a smooth function u_i on each Ω_i . Throughout this thesis, we denote the characteristic function (or indicator function) of Ω_i by

$$\chi_i(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \Omega_i, \\ 0, & \text{if } \mathbf{x} \in \Omega_i^c := \Omega \setminus \Omega_i, \end{cases} \quad (1.3)$$

which is a function of bounded variation in the following sense.

Definition 1.1 (TV and BV-Space). *Let Ω be an open domain in $\mathbb{R}^2$, and let u be a function in $L^1(\Omega)$. Then the total variation (TV) of u is defined by*

$$TV(u) = \int_{\Omega} |\nabla u| = \sup_{\mathbf{p} \in S} \int_{\Omega} u \operatorname{div} \mathbf{p} d\mathbf{x}, \quad (1.4)$$

where $\mathbf{p} = (p_1, p_2)$, $\operatorname{div} \mathbf{p} = \nabla \cdot \mathbf{p}$ and

$$S := \left\{ \mathbf{p}(\mathbf{x}) \in (C_c^1(\Omega))^2 : |\mathbf{p}(\mathbf{x})| = \sqrt{p_1(\mathbf{x})^2 + p_2(\mathbf{x})^2} \leq 1, \forall \mathbf{x} \in \Omega \right\}. \quad (1.5)$$

Here, $C_c^1(\Omega)$ is the space of continuously differentiable functions with compact support in Ω .

We define $BV(\Omega)$, the space of functions of bounded variation as

$$BV(\Omega) = \{u \in L^1(\Omega) : TV(u) < \infty\}, \quad (1.6)$$

with the norm

$$\|u\|_{BV} = \|u\|_{L^1} + TV(u).$$

It defines a Banach space. □

Note that if Ω is bounded, then the Sobolev space

$$W^{1,1}(\Omega) := \{u \in L^1(\Omega) : \nabla u \in L^1(\Omega)\} \subset BV(\Omega).$$

The following formula is centrally important:

$$TV(\chi_i) = \text{Perimeter}(\Omega_i) = |\partial\Omega_i|, \quad (1.7)$$

where χ_i is the characteristic function defined in (1.3). We refer to [52, 7, 47] for the properties of BV-space.

With the aid of (1.3) and (1.7), we formulate the Mumford-Shah functional as

$$\tilde{E}_{MS}(\mathbf{u}, \boldsymbol{\chi}) := \sum_{i=1}^m \left(\mu \int_{\Omega} |\nabla u_i|^2 \chi_i d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u_i - I|^2 \chi_i d\mathbf{x} + TV(\chi_i) \right), \quad (1.8)$$

where $\mathbf{u} = (u_1, \dots, u_m)$ and $\boldsymbol{\chi} = (\chi_1, \dots, \chi_m)$. Then it becomes natural to search for $\mathbf{u} \in (H^1(\Omega))^m$ ($H^1(\Omega)$ is the usual Sobolev space as defined in Adams [1]), and

$$\boldsymbol{\chi} \in \mathcal{A} := \left\{ \boldsymbol{\chi} \in (BV(\Omega))^m : \chi_i \in \{0, 1\}, \quad |\boldsymbol{\chi}| = \sum_{i=1}^m \chi_i = 1 \right\}. \quad (1.9)$$

Here, the constraints in $\mathcal{A}$ is the consequence of (1.2)-(1.3).

Remark 1.1. *The formulation (1.8) is actually a direct translation of the Mumford-Shah functional in (1.1) by using the characteristic functions. This naive form was firstly used in Vese and Chan [135], where the classical level-set method (cf. [98]) was used to approximate $\boldsymbol{\chi}$ and handle the related constraints (see Subsection 5.4.2 for a brief description of this algorithm).* □

We highlight that the minimization of (1.8) under (1.9) is challenging and worthy of deep investigation for the reasons:

- The admissible set $\mathcal{A}$ is nonconvex and the constraints are difficult to handle.
- It requires to resolve two vector unknowns, where the region to solve u_i is determined by χ_i , so they are intrinsically coupled together. Consequently, when the number of phases (i.e., m) is big, it is almost not possible to develop a robust minimization algorithm.
- We remark that in [135], a set of PDEs for $\{u_i\}$ has to be solved during the evolution of the level-set functions to obtain the smooth approximation of the given image I . The computational cost is expensive and the algorithm relies on the initialization.

With these issues in mind, the major motivations of this thesis are to propose new formulations to approximate the first regularization term in (1.1), and develop fast primal-dual minimization algorithms. We outline the main strategies and contributions along this line in Section 1.2 below.

1.1.2 Piecewise constant Mumford-Shah model

It is of interest to study the reduced Mumford-Shah model by dropping the first integral term of (1.1):

$$\min_{u, \Gamma} \left\{ E_{PCMS}(u, \Gamma) := \frac{\nu}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + |\Gamma| \right\}. \quad (1.10)$$

It is clear to see if Γ is fixed ($\Gamma \neq \emptyset$), then the solution u is piecewise constant (u is equal to the mean of I on the connected component of $\Omega \setminus \Gamma$), so E_{PCMS} becomes a function of Γ (cf. [47]). On the other hand, if u is a piecewise constant function, (1.1) reduces to (1.10). Hence, the model together with its level-set formulation is referred to as the piecewise constant Mumford-Shah (PCMS) model. Much work has been devoted to this particular case [92, 93, 94, 95, 84].

We particularly mention the Chan-Vese model [42] for two-phase image segmentation. Suppose that the given image I approximately consists of piecewise constant two phases of intensity value c_1 and c_2 on two regions separated by Γ . Denoting $\Omega_1 = \text{inside}(\Gamma)$ and $\Omega_2 = \text{outside}(\Gamma)$, the Chan-Vese (CV) model becomes

$$\min_{c_1, c_2, \Gamma} \left\{ E_{CV}(c_1, c_2, \Gamma) := \frac{\nu}{2} \left(\int_{\Omega_1} |c_1 - I|^2 d\mathbf{x} + \int_{\Omega_2} |c_2 - I|^2 d\mathbf{x} \right) + |\Gamma| \right\}. \quad (1.11)$$

The CV model, together with numerical implementation by the level-set method [98], has become a useful variational segmentation tool (see, e.g., [97, 127]).

It is straightforward to extend the CV model to multiphase piecewise constant Mumford-Shah model:

$$\min_{\mathbf{c}, \Gamma} \left\{ E_{PCMS}(\mathbf{c}, \Gamma) := \frac{\nu}{2} \sum_{i=1}^m \int_{\Omega_i} |c_i - I|^2 d\mathbf{x} + |\Gamma| \right\}, \quad (1.12)$$

where $\mathbf{c} = (c_1, \dots, c_m)$.

1.1.3 Potts model

As already mentioned, the multiclass labeling problem is closely related to the multiphase segmentation problem. In general, the multiclass labeling problem is to assign each pixel $\mathbf{x} \in \Omega$ a unique label l from a set of m labels (or classes) $\{l_1, \dots, l_m\}$. This problem can be formulated in the “discrete” or “continuous”, depending on the spatial definition of the labeling functions. However, it is NP-hard in the discrete context, as the number of unknowns grows exponentially with the number of labels. Hence, it becomes desirable to tackle the problem from the continuous perspective. Mathematically, the problem can be formulated by the (continuous) Potts model [108] (for multiclass labeling without favoring ordering):

$$\begin{aligned} \min_{\{\Omega_i\}_{i=1}^m} \left\{ E_{PTS}(\Gamma) := \lambda \sum_{i=1}^m \int_{\Omega_i} f_i(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^m |\partial\Omega_i| \right\}, \\ \text{subject to } \cup_{i=1}^m \Omega_i = \Omega \text{ with } \Omega_k \cap \Omega_l = \emptyset, \forall k \neq l, \end{aligned} \quad (1.13)$$

where Γ is the union of $\{\partial\Omega_i\}$ as before, the parameter $\lambda > 0$, and $\{f_i\}_{i=1}^m$ can be viewed as intrinsic forces to enforce the criterion of classification and labeling. It is clear that if $\mathbf{c}$ is assumed to be known a priori, then the PCMS model (1.12) turns out to be a special case of the Potts model (1.13).

1.1.4 Level-set methods

We see from the aforementioned models that their minimization requires to label and evolve $\{\Omega_i\}$, which actually can be realized using the versatile tool of the level set. Here, we review the essential ideas of the level-set method [98] and its important variant piecewise constant level-set method (PCLSM). Note that they differ in the way of labeling $\{\Omega_i\}$ and representing the characteristic functions $\{\chi_i\}$.

1. Level-set method by Osher and Sethian [98]

The level-set method, originally developed as numerical technique for tracking interfaces and shapes [98], has been increasingly applied on a broad spectrum of problems, including image analysis, reservoir simulations, inverse problems, computer vision and optimal shape design in the past decade [30, 42, 45, 45, 100, 81, 63, 26, 27]. Consider for example the partition of an open bounded domain $\Omega \in \mathbb{R}^2$ into two disjunctive subdomains by a closed curve Γ . The main idea is to embed Γ into a two-dimensional surface $z = \phi(x, y)$, and identify Γ as the zero level set: $\Gamma = \{\mathbf{x} = (x, y) \in \Omega : \phi(x, y) = 0\}$. A typical choice of ϕ is the signed distance function:

$$\phi(x, y) = \begin{cases} \text{dist}(\mathbf{x}, \Gamma), & \mathbf{x} \in \Omega_1 := \text{inside}(\Gamma), \\ -\text{dist}(\mathbf{x}, \Gamma), & \mathbf{x} \in \Omega_2 := \text{outside}(\Gamma). \end{cases} \quad (1.14)$$

Thus, we have

$$\Omega_1 = \{\mathbf{x} \in \Omega : \phi > 0\}, \quad \Omega_2 = \{\mathbf{x} \in \Omega : \phi < 0\}.$$

Using the one-dimensional Heaviside function

$$H(\phi) = \begin{cases} 0, & \text{if } \phi \leq 0, \\ 1, & \text{if } \phi > 0, \end{cases} \quad (1.15)$$

we can express the characteristic functions as

$$\chi_1 = H(\phi), \quad \chi_2 = 1 - H(\phi). \quad (1.16)$$

We find from (1.7) and the fact $\delta(\phi) = H'(\phi)$ (where δ is the Dirac delta function) that (see [97, (1.21)]):

$$|\Gamma| = TV(H(\phi)) = \int_{\Omega} \delta(\phi) |\nabla \phi| d\mathbf{x}. \quad (1.17)$$

One also verifies that the signed distance function satisfies the Eikron equation

$$|\nabla \phi| = 1. \quad (1.18)$$

Remark 1.2. *Equipped with the level-set function, one can also represent other geometric quantities, e.g., area and curvature (cf. [97]).* $\square$

This notion can be extended to multidomains by using multiple level-set functions [135]. For example, we consider the partition of Ω into four disjoint subdomains $\{\Omega_i\}_{i=1}^4$. Using the sign of signed distance functions ϕ_1 and ϕ_2 , the curves $\{\phi_1 = 0\} \cup \{\phi_2 = 0\}$ can determine four subdomains (see Figure 1.1):

$$\begin{aligned} \Omega_1 &= \{\mathbf{x} \in \Omega : \phi_1 > 0, \phi_2 > 0\}, & \Omega_2 &= \{\mathbf{x} \in \Omega : \phi_1 > 0, \phi_2 < 0\}, \\ \Omega_3 &= \{\mathbf{x} \in \Omega : \phi_1 < 0, \phi_2 > 0\}, & \Omega_4 &= \{\mathbf{x} \in \Omega : \phi_1 < 0, \phi_2 < 0\}. \end{aligned} \quad (1.19)$$

In general, m level-set functions can label 2^m subdomains. Similarly, we can express the characteristic functions, e.g., as

$$\chi_1 = H(\phi_1)H(\phi_2), \quad \chi_2 = H(\phi_1)(1 - H(\phi_2)).$$

Armed with the level-set function, we can write the level set formulation of the CV functional in (1.11) as

$$\begin{aligned} E_{CV}(c_1, c_2, \phi) &:= \frac{\nu}{2} \left(\int_{\Omega} |c_1 - I|^2 H(\phi) d\mathbf{x} + \int_{\Omega} |c_2 - I|^2 (1 - H(\phi)) d\mathbf{x} \right) \\ &\quad + \int_{\Omega} \delta(\phi) |\nabla \phi| d\mathbf{x}. \end{aligned} \quad (1.20)$$

Figure 1.1: Labeling of four subdomains by two signed distance functions.

In real implementation, we use the smeared-out Heaviside function and delta function to replace H and δ (see [97, 42, 12]):

$$H_\epsilon(x) = \frac{1}{2} \left(1 + \frac{2}{\pi} \arctan \frac{x}{\epsilon} \right), \quad \delta_\epsilon(x) = H'_\epsilon(x) = \frac{\epsilon}{\pi(\epsilon^2 + x^2)}. \quad (1.21)$$

Keeping ϕ fixed and minimizing (1.20) with respect to c_1 and c_2 , we have

$$c_1 = \frac{\int_\Omega I H_\epsilon(\phi) d\mathbf{x}}{\int_\Omega H_\epsilon(\phi) d\mathbf{x}}, \quad c_2 = \frac{\int_\Omega I(1 - H_\epsilon(\phi)) d\mathbf{x}}{\int_\Omega (1 - H_\epsilon(\phi)) d\mathbf{x}}. \quad (1.22)$$

With fixed c_1 and c_2 , minimizing (1.20) with respect to ϕ yields the following Euler-Lagrange equation:

$$\delta_\epsilon(\phi) \left[\operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \nu(c_1 - I)^2 + \nu(c_2 - I)^2 \right] = 0. \quad (1.23)$$

In the implementation, an appropriate re-initialization is needed to enforce ϕ_j to stay close to a signed distance function [101, 114, 97]. It is interesting to point out that in [67] (for two-phase segmentation), the constraint (1.18) is built in the variational formulation by penalizing the term $\int_\Omega (|\nabla \phi|^2 - 1)^2 d\mathbf{x}$, where ϕ is the targeted signed distance function.

2. Piecewise constant level-set method

As an important variance, the binary level-set method proposed by [72] identifies the contour Γ as the discontinuity of the piecewise constant level-set function (see Figure 1.2 (left)):

$$\psi(x, y) = \begin{cases} 1, & \mathbf{x} \in \Omega_1 = \text{inside}(\Gamma), \\ -1, & \mathbf{x} \in \Omega_2 = \text{outside}(\Gamma). \end{cases} \quad (1.24)$$

In contrast to (1.16), the characteristic functions

$$\chi_1 = \frac{1 + \psi}{2}, \quad \chi_2 = \frac{1 - \psi}{2}, \quad (1.25)$$

and the counterpart of (1.17) becomes

$$|\Gamma| = TV(\chi_1) = \frac{1}{2} \int_{\Omega} |\nabla \psi| d\mathbf{x}. \quad (1.26)$$

In this case, the formulation of the CV functional in (1.11) becomes

$$\begin{aligned} E_{CV}(c_1, c_2, \psi) := & \frac{\nu}{2} \left(\int_{\Omega} |c_1 - I|^2 \frac{1 + \psi}{2} d\mathbf{x} + \int_{\Omega} |c_2 - I|^2 \frac{1 - \psi}{2} d\mathbf{x} \right) \\ & + \int_{\Omega} |\nabla \psi| d\mathbf{x}, \end{aligned} \quad (1.27)$$

where we impose the constraint $\psi^2 = 1$ to ensure (1.24). In Chapter 2, we deal with the constraint by a penalization technique, and present the associated gradient descent PDE, which is termed as the TV-Allen-Cahn model, and provide an in-depth study of this model.

Remark 1.3. We see from (1.15)-(1.16) and (1.24)-(1.25) that

$$\psi = 2H(\phi) - 1,$$

which indicates that ψ is a composite of H and the signed distance function ϕ . $\square$

Remark 1.4. In [72, 73], the augmented Lagrangian technique was used to deal with the constraint $\phi^2 = 1$. $\square$

Figure 1.2: Signed distance function (left) versus piecewise constant level-set function (right) for representing a circle

The binary level-set method was further extended to multiphase problem in [72, 73]. Basically, to partition Ω into m subdomains as in (1.2), we label each subdomain by an integer and define the piecewise constant level-set function as

$$\psi(\mathbf{x}) = i, \quad \forall \mathbf{x} \in \Omega_i, \quad \text{for all } i = 1, 2, \dots, m. \quad (1.28)$$

The characteristic function of Ω_i is given by the Lagrange basis polynomial:

$$\chi_i = \frac{1}{\alpha_i} \prod_{j=1, j \neq i}^m (\psi - j) \quad \text{and} \quad \alpha_i = \prod_{k=1, k \neq i}^m (i - k). \quad (1.29)$$

Thus, the length of $\partial\Omega_i$ can be represented as

$$|\partial\Omega_i| = TV(\chi_i) = \int_{\Omega} |\nabla \chi_i| d\mathbf{x}, \quad 1 \leq i \leq m. \quad (1.30)$$

In this situation, the following constraint is indispensable

$$K_m(\psi) = (\psi - 1)(\psi - 2) \cdots (\psi - m) = 0. \quad (1.31)$$

Then the approximation of the given piecewise constant image I is given by

$$I \approx u = c_1 \chi_1 + c_2 \chi_2 + \cdots + c_m \chi_m. \quad (1.32)$$

Based on the above observations, we reformulate the multiphase piecewise constant model (1.12) to be

$$\min_{\mathbf{c}, \psi, K_m(\psi)=0} \left\{ E_{PCMS}(\mathbf{c}, \psi) := \frac{\nu}{2} \sum_{i=1}^m \int_{\Omega} |c_i - I|^2 \chi_i d\mathbf{x} + \sum_{i=1}^m \int_{\Omega} |\nabla \chi_i| d\mathbf{x} \right\}. \quad (1.33)$$

Here, with the constraint $K_m(\psi) = 0$, we use the penalization method, so the corresponding model becomes

$$\min_{\mathbf{c}, \psi} \left\{ E_{PCMS}(\mathbf{c}, \psi) := \frac{\nu}{2} \sum_{i=1}^m \int_{\Omega} |c_i - I|^2 \chi_i d\mathbf{x} + \sum_{i=1}^m \int_{\Omega} |\nabla \chi_i| d\mathbf{x} + \frac{r}{2} \int_{\Omega} |K_m(\psi)|^2 d\mathbf{x} \right\}, \quad (1.34)$$

where $r > 0$ is the constant. Then the Euler-Lagrange equation can be formulated as:

$$\frac{\partial \psi}{\partial t} = \sum_{i=1}^m \operatorname{div} \left(\frac{\nabla \chi_i}{|\nabla \chi_i|} \right) \frac{\partial \chi_i}{\partial \psi} - \frac{\nu}{2} \sum_{i=1}^m (c_i - I)^2 \frac{\partial \chi_i}{\partial \psi} - r K_m(\psi) K'_m(\psi). \quad (1.35)$$

For a given ψ , the phase values $\{c_i\}$ are determined by solving the linear system:

$$\sum_{j=1}^m c_j \int_{\Omega} \chi_i \chi_j d\mathbf{x} = \int_{\Omega} I \chi_i d\mathbf{x}, \quad \text{for } i = 1, \dots, m. \quad (1.36)$$

Remark 1.5. *The PCLSM has been applied to the multiphase image segmentation [73, 110], and other interface problems [71, 79]. Compared with the level-set methods using signed distance functions, this only uses one level-set function to represent multiple phases. Some advantages are listed in [72, 73].*

It should be pointed out that the essential tool (1.28)-(1.29) results from a global polynomial interpolation, so it requires to manipulate polynomials of high degree if m is large. Therefore, the stability of the algorithm is a severe issue in this context.

- *The leading term of the Euler-Lagrange equation of ϕ is of the form:*

$$- \sum_{i=1}^m \operatorname{div} \left(\frac{\nabla \chi_i}{|\nabla \chi_i|} \right) \frac{d\chi_i}{d\psi}, \quad (1.37)$$

so the equation to evolve ψ is highly nonlinear, and degenerating at the zeros of the “polynomial” $\frac{d\chi_i}{d\psi}$. One may replace the second term of (1.34) by $TV(\psi)$. This seems to be a big modification, but the numerical evidences (see, e.g., [110]) show that this model produces quite satisfactory segmentation.

- Since the subdomains are labeled by the zeros of the global polynomial K_m , a small perturbation of labels leads to a big error of the values of K_m and K'_m , or a misidentification of the subdomains, when the number of phases is slightly large. This significantly requires a good input of the initial values and a careful selection of the parameters. This suggests a resort of a stable lower-order piecewise interpolation to label the phases. $\square$

3. Level-set method using characteristic functions

In view of the drawbacks of the foregoing level-set methods, we advocate the direct use of the characteristic functions (cf. (1.9)) for labeling each subdomain. Hereafter, we denote

$$v_i(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in \Omega_i, \\ 0, & \text{if } \mathbf{x} \in \Omega_i^c := \Omega \setminus \Omega_i, \end{cases} \quad (1.38)$$

for all $1 \leq i \leq m$. Then we have

$$|\partial\Omega_i| = TV(v_i) = \int_{\Omega} |\nabla v_i| d\mathbf{x}, \quad 1 \leq i \leq m, \quad (1.39)$$

and the associated constraints become

$$v_i \in \{0, 1\}, \quad 1 \leq i \leq m; \quad \sum_{i=1}^m v_i = 1. \quad (1.40)$$

Hence, the approximation of the given piecewise constant image I is given by

$$I \approx u = c_1 v_1 + c_2 v_2 + \cdots + c_m v_m. \quad (1.41)$$

Remark 1.6. Note that for $m = 2$, the constraint (1.40) implies $v_2 = 1 - v_1$, so we can eliminate v_2 . This will lead the binary level-set method as in [72]. More detailed discussions will be given in Chapter 3.

As a distinctive difference, the approximation (1.41) is actually a piecewise constant interpolation of the intensity values, while (1.32) is a global polynomial interpolation. Hence, it is anticipated that the approach based on (1.41) is more stable for large m . $\square$

Under this notion, we can formulate the models by directly using $\{v_i\}_{i=1}^m$. More precisely, like (1.33), we have

$$\min_{\mathbf{c}, \mathbf{v} \in \mathcal{A}} \left\{ E_{PCMS}(\mathbf{c}, \mathbf{v}) := \frac{\nu}{2} \sum_{i=1}^m \int_{\Omega} |c_i - I|^2 v_i d\mathbf{x} + \sum_{i=1}^m \int_{\Omega} |\nabla v_i| d\mathbf{x} \right\}, \quad (1.42)$$

where $\mathcal{A}$ is defined in (1.9). As mentioned before, the set $\mathcal{A}$ is nonconvex. The question of how to construct a fast global minimization algorithm becomes centrally important, and this will be the main topic of Chapter 3.

It should be pointed out that in addition to the above background knowledge, some other aspects in image processing, such as the Rudin-Osher- Fatemi (ROF) model [112], the Chambolle's dual algorithm [34], and the proximal algorithm [109], will be properly reviewed in the forthcoming chapters.

1.2 Main Contributions of the Dissertation

This thesis aims to propose new modified Mumford-Shah models, and develop fast numerical algorithms for image segmentation and multiclass labeling problems. The main contents and contributions of this dissertation are highlighted as follows.

- In Chapter 2, we focus on two-phase image segmentation by using the binary level-set method. The direct use of a penalization technique to treat the constraint $\psi^2 = 1$ leads to a fully nonlinear TV-Allen-Cahn model. The commonly-used numerical approaches usually suffer from not only the non-differentiability of the TV-term, but also directly evolving the discontinuous piecewise constant-structured solutions. In this chapter, we propose efficient dual algorithms to overcome these difficulties. The use of a splitting-penalty method results in TV-Allen-Cahn type models associated with different “double-well” potentials, which allow for the implementation of the dual algorithm of Chambolle [34]. Moreover, we present a new dual algorithm

based on an edge-featured penalty of the dual variable, which only requires to solve a vectorial Allen-Cahn type equation with linear $\nabla(\text{div})$ -diffusion rather than fully nonlinear diffusion in the Chambolle's approach. Consequently, more efficient numerical algorithms such as time-splitting method and Fast Fourier Transform (FFT) can be implemented. Various numerical tests show that two dual algorithms are much faster and more stable than the primal gradient descent approach, and the new dual algorithm is at least as efficient as the Chambolle's algorithm and is relatively accurate. We demonstrate that the new method also provides a viable alternative for image restoration.

- In Chapter 3, we develop fast global minimization algorithm for the Potts model (1.13) and the multiphase piecewise constant Mumford-Shah model (1.42). Different from the existing methods, we work directly with the binary setting without using convex relaxation, which is termed as a direct approach. We provide the sufficient and necessary conditions to guarantee a global optimum. Moreover, we develop efficient algorithms based on a reduction of the intermediate unknowns from the augmented Lagrangian formulation. As a result, the underlying algorithms involve significantly fewer parameters and unknowns than the naive use of augmented Lagrangian-based methods, so they are fast and easy to implement. Furthermore, they can produce a global optimum under mild conditions.
- In Chapter 4, we propose efficient algorithms for the modified Mumford-Shah model in [137] (which immerses an edge into a narrow band surrounding it, and measures the edge length by TV of the associated binary level-set function). Limited algorithms were proposed due to the situation that the edge set could be of measure zero in the limiting process. In order to overcome the difficulty, the relaxation technique and splitting-penalty

method are conducted to manage the constraint of binary level-set function, and then we introduce two fast algorithms using the proximity algorithm and split Bregman method for solving the subproblem of the level-set function, which can mathematically handle the limiting case. Moreover, this approach can also be extended to color image processing. Finally, numerical results are provided to show the efficiency and robustness of the proposed algorithms.

- In Chapter 5, we develop a decomposition model for piecewise smooth image segmentation. By decomposing an image into an interpolation of piecewise constant part, smooth component and noise part, we incorporate these components into the Mumford-Shah model. This only requires to solve the smooth component into the Mumford-Shah model. Then the relaxation technique and analysis of existence are presented. With the entropy penalization, we construct efficient algorithms which include the Fourier method, smoothed Chambolle-Dual algorithm and proximity algorithm. Moreover, the numerical results show the advantages of the proposed algorithms over the existing method.
- The last chapter is for some concluding remarks and future works.

Efficient Dual Algorithms for TV-Allen-Cahn Model

This chapter is devoted to fast dual algorithms for two-phase image segmentation using the reduced Mumford-Shah model (1.10). As shown in Chapter 1, using the piecewise constant level-set method (cf. (1.24)) leads to the model (cf. [72]):

$$\begin{cases} \min_{c_1, c_2, \psi} \left\{ TV(\psi) + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} \right\}, & \lambda > 0, \\ \text{subject to } \psi^2 = 1, \end{cases} \quad (2.1)$$

where I is the given image defined on a bounded open domain in $\mathbb{R}^2$, ψ is the level-set function, and

$$u = c_1 \frac{1 + \psi}{2} + c_2 \frac{1 - \psi}{2} \approx I. \quad (2.2)$$

The constraint is imposed to ensure there is no vacuum and overlapping between different phases, that is, each point $\mathbf{x} \in \Omega$ can take one and only one phase value. We refer to Chapter 1 for the definition of total variation (TV).

In [73], the constraint was treated by the augmented Lagrangian technique. Here, we use the penalty method, so the constrained minimization problem (2.1) can be approximated by (cf. [72]):

$$\min_{c_1, c_2, \psi} \left\{ TV(\psi) + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + \frac{\mu}{4} \int_{\Omega} (\psi^2 - 1)^2 d\mathbf{x} \right\}, \quad \lambda, \mu > 0, \quad (2.3)$$

where u is given by (2.2). The optimality conditions lead to

$$-\operatorname{div} \left(\frac{\nabla \psi}{|\nabla \psi|} \right) + \lambda(u - I) \frac{\partial u}{\partial \psi} + \mu \psi (\psi^2 - 1) = 0, \quad (2.4)$$

and

$$\int_{\Omega} u \frac{\partial u}{\partial c_i} d\mathbf{x} - \int_{\Omega} I \frac{\partial u}{\partial c_i} d\mathbf{x} = 0, \quad i = 1, 2. \quad (2.5)$$

For fixed ψ , the approximation u and the phase intensities c_1, c_2 can be updated by (2.2) and by solving the simple system (2.5), respectively. Hence, the effort needs to be devoted to the nonlinear stiff PDE (2.4) with fixed c_1 and c_2 . Most of the existing methods are based on evolving the gradient descent flow

$$\frac{\partial \psi}{\partial t} = \operatorname{div} \left(\frac{\nabla \psi}{|\nabla \psi|} \right) - \lambda(u - I) \frac{\partial u}{\partial \psi} + \mu(1 - \psi^2)\psi, \quad (2.6)$$

which is closely related to the Allen-Cahn model for phase transitions [3] but with a nonlinear TV-diffusion operator rather than a linear Laplace-diffusion. To distinguish them, we term it as the TV-Allen-Cahn model. The presence of nonlinear diffusion induces some numerical difficulties, but since the TV-term has the capability to preserve discontinuities and edges, the model (2.6) might avoid mesh refinement as usually required for phase-field models [55, 119]. Some algorithms have been developed and tested for numerical solution of (2.6), see e.g., [72, 73, 71, 79, 38, 43]. However, direct evolution of the primal variable ψ might suffer from numerical difficulties due to the nonlinearity and non-differentiability of $TV(\psi)$ in (2.3). Note that the denominator of the TV-term has to be computed approximately by a regularized one: $|\nabla \psi|_{\beta} = \sqrt{\psi_x^2 + \psi_y^2 + \beta}$ for some $\beta > 0$. In addition, time evolution of piecewise-constant-structured solutions inevitably leads to a much severer restriction to time step.

The dual algorithms have been proven to be efficient to overcome these difficulties. So far, much progress has been made in dual methods for the ROF model [40, 34]. Most notably, Chambolle [34] provided a convergence analysis for the proposed dual algorithm. In addition, by introducing the auxiliary variable: $\mathbf{q} = \nabla \psi$ to handle the TV-term (see (2.25) below), splitting-penalty and/or augmented Lagrangian type methods have been developed in [49, 138, 61, 129] for

the ROF model. Interestingly, using the augmented Lagrangian method, Tai and Wu [129] presented a uniform framework to derive the dual algorithms [40, 34] and splitting-penalty based methods [49, 138, 61].

Motivated by the success of dual algorithms for image restoration [40, 34, 61], we propose in this chapter two efficient dual algorithms to overcome the aforementioned drawbacks of the primal approaches. The key idea is to introduce an auxiliary variable and use a splitting-penalty method, which results in a minimization problem of Rudin-Osher-Fatemi (ROF) [112] type. Accordingly, this forms the base to implement efficient dual algorithms. It is worthwhile to point out that the splitting-penalty approach has been around as early as Courant [44] in 1943, and has been applied in various context of image restoration [49, 138, 61, 129]. Indeed, it has proven to be an effective tool for convex optimization. However, we realize that for this application, the splitting-penalty technique leads to TV-Allen-Cahn type models associated with “double-well” potentials different from the one (i.e., $(\psi^2 - 1)^2$ with ψ being the phase function) in the original Allen-Cahn equation, so the penalty parameter is not necessary to be sufficiently large (slightly bigger than the parameter μ in (2.7) below), which should be in contrast to the usual penalty approach. In this chapter, we first extend the Chambolle’s dual algorithm [34] to such TV-Allen-Cahn type models. More importantly, we propose a new dual algorithm based on an edge-featured penalty of the dual variable, which only requires to solve a vectorial Allen-Cahn type models with linear $\nabla(\text{div})$ -diffusion rather than fully nonlinear diffusion in the Chambolle’s approach. Consequently, more efficient algorithms can be implemented on the new model such as an operator-splitting in time and FFT in space as suggested in this chapter.

The rest of the chapter is organized as follows. In Section 2.1, we apply a splitting-penalty method to obtain TV-Allen-Cahn-type model in primal variable with various potentials. We demonstrate that the models produce desirable segmentation results as the TV-Allen-Cahn equation associated with the usual

“double-well” potential. We present in Sections 2.2 and 2.3 the dual algorithms based on Chambolle’s idea, and the new dual algorithm using a penalty of the dual variable along edges and implemented by an operator splitting in time and FFT in space. In Section 2.4, we test the algorithms on various images, and demonstrate the dual algorithms are much faster than the primal approaches, and the new algorithm slightly outperforms the dual algorithm based on Chambolle’s approach in most experiments. We conclude the chapter with some discussions on the applications of the new algorithm to image restoration.

2.1 Splitting and Penalty Technique

Our starting point is to introduce an auxiliary variable and rewrite (2.3) into

$$\min_{c_1, c_2, \psi, v} \left\{ TV(\psi) + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + \frac{\mu}{4} \int_{\Omega} (v^2 - 1)^2 d\mathbf{x} \right\} \quad \text{subject to } v = \psi. \quad (2.7)$$

This constrained problem can be approximated by penalizing the constraint:

$$\min_{c_1, c_2, \psi, v} \left\{ TV(\psi) + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + \frac{\mu}{4} \int_{\Omega} (v^2 - 1)^2 d\mathbf{x} + \frac{\theta}{2} \int_{\Omega} (v - \psi)^2 d\mathbf{x} \right\}, \quad (2.8)$$

where the penalty parameter $\theta > 0$. It is clear that if θ is sufficiently large, the problem (2.8) can accurately approximate (2.3). However, we next show that for $\theta \geq \mu$, the model (2.8) is also robust for segmentation, which leads to TV-Allen-Cahn-type model involving a potential different from the usual “double-well” one. As for the splitting technique in [22], we seek the minimizer (u, v, c_1, c_2) separately as follows.

(a). Fixing v, c_1 and c_2 , we search for ψ as a solution of

$$\min_{\psi} \left\{ TV(\psi) + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + \frac{\theta}{2} \int_{\Omega} (v - \psi)^2 d\mathbf{x} \right\}. \quad (2.9)$$

(b). Fixing ψ, c_1 and c_2 , we search for v as a solution of

$$\min_v \left\{ \frac{\mu}{4} \int_{\Omega} (v^2 - 1)^2 d\mathbf{x} + \frac{\theta}{2} \int_{\Omega} (v - \psi)^2 d\mathbf{x} \right\}. \quad (2.10)$$

(c). Fixing ψ and v , the approximation (c_1, c_2) of the intensities can be computed by (2.5).

It is clear that the gradient descent flow of (2.9) takes the form

$$\frac{\partial \psi}{\partial t} = \operatorname{div} \left(\frac{\nabla \psi}{|\nabla \psi|} \right) - \lambda(u - I) \frac{\partial u}{\partial \psi} + \theta(v - \psi), \quad \text{in } \Omega, \quad (2.11)$$

and the optimality condition of (2.10) implies

$$v^3 + (r - 1)v = r\psi \quad \text{with} \quad r = \frac{\theta}{\mu}. \quad (2.12)$$

Hereafter, we assume that $r \geq 1$, and find that the unique real root of (2.12) is

$$v = G_r(\psi) := \sqrt[3]{\frac{r\psi}{2} + \sqrt{\frac{r^2\psi^2}{4} + \frac{(r-1)^3}{27}}} + \sqrt[3]{\frac{r\psi}{2} - \sqrt{\frac{r^2\psi^2}{4} + \frac{(r-1)^3}{27}}}. \quad (2.13)$$

It is apparent that

$$r(G_r(\psi) - \psi) = G_r(\psi) - G_r^3(\psi). \quad (2.14)$$

The equation (2.11) can be reformulated as

$$\frac{\partial \psi}{\partial t} = \operatorname{div} \left(\frac{\nabla \psi}{|\nabla \psi|} \right) - \lambda(u - I) \frac{\partial u}{\partial \psi} + \mu(1 - G_r^2(\psi))G_r(\psi), \quad 1 \leq r \leq \infty. \quad (2.15)$$

It can be regarded as a modified TV-Allen-Cahn model associated with the “double-well” potential:

$$W_r(\psi) = \int_{-1}^{\psi} (1 - G_r^2(\varphi))G_r(\varphi)d\varphi, \quad 1 \leq r \leq \infty, \quad (2.16)$$

in contrast to the usual “double-well” potential: $W(\psi) = (\psi^2 - 1)^2/4$ corresponding to the TV-Allen-Cahn model (2.6). From a slightly different point of view, the equation (2.15) is the gradient descent flow of (2.1) but with a different constraint to the piecewise constant level-set function ψ , and accordingly with the penalty term $\mu \int_{\Omega} W_r(\psi)d\mathbf{x}$ in place of $\mu \int_{\Omega} W(\psi)d\mathbf{x}$ in (2.3).

One verifies that

$$G_1(\psi) = \sqrt[3]{\psi}; \quad G_r(\psi) = \psi + O(r^{-3/2}), \quad r \gg 1, \quad (2.17)$$

which implies

$$\lim_{r \rightarrow \infty} \left\{ (1 - G_r^2(\varphi)) G_r(\varphi) \right\} = (1 - \psi^2)\psi; \quad \lim_{r \rightarrow \infty} W_r(\psi) = W(\psi). \quad (2.18)$$

In addition, we find from (2.12) that

$$\frac{dG_r}{d\psi} = \frac{r}{r - 1 + 3G_r^2(\psi)},$$

which indicates that for $r > 1$, $G_r(\psi)$ is sufficiently smooth, so is $W_r(\psi)$. In Figure 2.1(a), we plot the graphs of several samples of $W_r(\psi)$ with $\psi \in [-1.5, 1.5]$ and for $r = 1, 2, 5, 10, 50, 300$ from bottom to top by solid lines, and plot the graph of $W(\psi)$ marked by “o”. We see that the “double wells” with $r > 1$ are smooth, and $W_{300}(\psi)$ is almost indistinguishable from $W(\psi) = W_\infty(\psi)$.

To illustrate the performance of the TV-Allen-Cahn type models (2.15) in terms of the choice of r , we discretize it by an explicit scheme:

$$\frac{\psi^{n+1} - \psi^n}{\tau} = \operatorname{div} \left(\frac{\nabla \psi^n}{|\nabla \psi^n|_\beta} \right) - \lambda(u^n - I) \frac{c_1^n - c_2^n}{2} + \mu(1 - G_r^2(\psi^n)) G_r(\psi^n), \quad (2.19)$$

where $|\nabla \psi^n|_\beta = \sqrt{(\psi_x^n)^2 + (\psi_y^n)^2 + \beta}$ for some small $\beta > 0$. The equation is equipped with the Neumann boundary condition: $\partial\psi/\partial\mathbf{n} = 0$ as usual, and the divergence term is discretized by a central difference scheme as in [97] (cf. P. 103). Note that given ψ^n , the values of u^n , c_1^n and c_2^n are evaluated by (2.2) and (2.5). We use the annulus-square image in Figure 2.1(d) as an example, and choose the parameters: $\lambda = 1$, $\mu = 200$, $\tau = 5e - 5$ and $\beta = 1e - 6$. We take the initial value of ψ to be 2 inside the red box, and -2 outside (see Figure 2.1(d)). To measure the convergence and accuracy, we define pixel-wise errors

$$\begin{aligned} E_{\max}\{\psi^{n+1}, \psi^n\} &= \max_{i,j} |\psi^{n+1} - \psi^n|, \\ E\{|\psi^n|, 1\} &= \max \left\{ \left| \max_{i,j} |\psi^n| - 1 \right|, \left| \min_{i,j} |\psi^n| - 1 \right| \right\}. \end{aligned} \quad (2.20)$$

In Figure 2.1(b) and 2.1(c), we record the history of $E_{\max}\{\psi^{n+1}, \psi^n\}$ and the errors $E\{|\psi^n|, 1\}$ to the targeted piecewise constant function (1.24). We observe

Figure 2.1: Profiles of the “double-well” potentials, convergence behavior (from right to left: $r = 1, 2, \infty, 5, 10, 50, 300$), and segmented results obtained by (2.19).

a similar convergence behavior for all cases except for $r = 1$. Notice that in case of $r = 1$, the potential $W_r(\psi)$ is not differentiable at $\psi = 0$. We plot in Figure 2.1(f) the profile of the phase function ψ at 180 steps, which shows a good approximation of the piecewise constant level-set as expected. In Figure 2.1(e), we depict the boundaries after segmentation, which indicates a quite accurate detection for $r = 10$ at 180 steps.

As with ROF model based image restoration, the primal approach is born with slow convergence and severe restriction on time stepping size, although some improvement can be made by using, e.g., the AOS scheme [77, 78, 140]. Moreover, in this context, it is also quite sensitive to the choice of ψ^0 (refer to the second numerical example in Section 2.4). We next introduce two dual algorithms to overcome these shortcomings.

2.2 Dual Algorithm based on Chambolle [34]

Now, we are in a position to present the fast dual algorithms for the minimization problem (2.8). Since the subproblem (2.10) can be solved analytically, we only need to focus on the subproblem (2.9).

As u is a linear function of ψ , the subproblem (2.9) with fixed c_1, c_2 and v can be solved by using Chambolle's dual algorithm. More precisely, the level-set function ψ is computed by

$$\begin{aligned} \psi &= \mathcal{C}(\mathbf{p}, v, c_1, c_2) \\ &:= \frac{1}{\lambda(c_1 - c_2)^2/4 + \theta} \left(\frac{\lambda(c_1 - c_2)}{2} I - \operatorname{div} \mathbf{p} + \theta v - \lambda \frac{c_1^2 - c_2^2}{4} \right), \end{aligned} \quad (2.21)$$

and the dual variable $\mathbf{p} = (p_1, p_2)$ satisfies

$$-\mathbf{H} + |\mathbf{H}| \mathbf{p} = 0, \quad (2.22)$$

where

$$\mathbf{H} := \mathbf{H}(\mathbf{p}, v, c_1, c_2) = \nabla \left(\operatorname{div} \mathbf{p} - \frac{\lambda(c_1 - c_2)}{2} I - \theta v \right). \quad (2.23)$$

For fixed c_1, c_2 and v , the problem (2.22)-(2.23) can be solved by the iteration with $\mathbf{p}^0 = 0$:

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \tau \mathbf{H}(\mathbf{p}^n, v, c_1, c_2)}{1 + \tau |\mathbf{H}(\mathbf{p}^n, v, c_1, c_2)|}, \quad n = 0, 1, \dots \quad (2.24)$$

It is anticipated that the iteration converges for $0 < \tau \leq \frac{1}{8}$ as with the original algorithm for ROF model in [36], and might be sped up by using the technique in [35]. In addition, one may follow [36] to derive (2.22) by using the Karush-Kuhn-Tucker conditions for inequality constrained optimization problems. However, it can be derived by a much simpler argument as in [129]. For the interested readers' reference, we briefly sketch the derivation followed from [129] below. We first convert the subproblem (2.9) into the following constrained minimization:

$$\begin{cases} \min_{\mathbf{q}, \psi} \left\{ \int_{\Omega} |\mathbf{q}| d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + \frac{\theta}{2} \int_{\Omega} (v - \psi)^2 d\mathbf{x} \right\}, \\ \text{subject to } \mathbf{q} = \nabla \psi. \end{cases} \quad (2.25)$$

Using the notion of augmented Lagrangian method, we obtain

$$\begin{aligned} \min_{\psi, \mathbf{q}} \max_{\mathbf{p}} \left\{ \mathcal{F}(\psi, \mathbf{q}, \mathbf{p}) := \int_{\Omega} |\mathbf{q}| d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} \right. \\ \left. + \frac{\theta}{2} \int_{\Omega} (v - \psi)^2 d\mathbf{x} + \int_{\Omega} \mathbf{p} \cdot (\mathbf{q} - \nabla \psi) d\mathbf{x} \right. \\ \left. + \frac{\gamma}{2} \int_{\Omega} |\mathbf{q} - \nabla \psi|^2 d\mathbf{x} \right\}, \quad \lambda, \gamma, \theta > 0, \end{aligned} \quad (2.26)$$

where $\mathbf{p} = (p_1, p_2)$ is the Lagrange multiplier. By the optimality conditions,

$$\frac{\partial \mathcal{F}(\psi, \mathbf{q}, \mathbf{p})}{\partial \psi} = \lambda(u - I) \frac{\partial u}{\partial \psi} - \theta(v - \psi) + \operatorname{div} \mathbf{p} + \gamma \operatorname{div}(\mathbf{q} - \nabla \psi) = 0, \quad (2.27a)$$

$$\frac{\partial \mathcal{F}(\psi, \mathbf{q}, \mathbf{p})}{\partial \mathbf{q}} = \frac{\mathbf{q}}{|\mathbf{q}|} + \mathbf{p} + \gamma(\mathbf{q} - \nabla \psi) = 0, \quad (2.27b)$$

$$\frac{\partial \mathcal{F}(\psi, \mathbf{q}, \mathbf{p})}{\partial \mathbf{p}} = \mathbf{q} - \nabla \psi = 0. \quad (2.27c)$$

we find from (2.27a) and (2.27c) that

$$\lambda(u - I) \frac{\partial u}{\partial \psi} - \theta(v - \psi) + \operatorname{div} \mathbf{p} = 0. \quad (2.28)$$

In view of (2.2), we can express ψ in terms of the dual variable $\mathbf{p}$ and obtain (2.21). Moreover, the equation (2.22) follows from (2.27b) and (2.27c). Finally, scaling out the constant factor

$$d := d(c_1, c_2) = \frac{\lambda}{4}(c_1 - c_2)^2 + \theta, \quad (2.29)$$

and applying a semi-implicit scheme to the corresponding gradient descent flow (cf. [34]), we obtain the scheme (2.24).

As usual, a homogeneous Dirichlet boundary condition: $\mathbf{p}|_{\partial\Omega} = 0$ can be imposed. Here, we use a periodic boundary condition for comparison with the new dual algorithm in the next section. We also note that $\mathbf{H}$ can be discretized by finite difference as in [34].

We summarize the dual algorithm for (2.9)-(2.10) as Algorithm 1 below.

Algorithm 1.

1. Initialization: choose ψ^0 and set $\mathbf{p}^0 = 0$.
2. For $n = 0, 1, \dots$,
 - (i) Compute (c_1^n, c_2^n) by solving the simple system (2.5) with $\psi = \psi^n$.
 - (ii) Compute u^n by

$$u^n = \frac{1 + \psi^n}{2} c_1^n + \frac{1 - \psi^n}{2} c_2^n.$$
 - (iii) Compute v^n by $v^n = G_r(\psi^n)$ in (2.13).
 - (iv) Compute $\mathbf{p}^{n+1}$ by (2.24):

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \tau \mathbf{H}(\mathbf{p}^n, v^n, c_1^n, c_2^n)}{1 + \tau |\mathbf{H}(\mathbf{p}^n, v^n, c_1^n, c_2^n)|}.$$
 - (v) Compute ψ^{n+1} by (2.21):

$$\psi^{n+1} = \mathcal{C}(\mathbf{p}^{n+1}, v^n, c_1^n, c_2^n).$$
3. Goto 2 till some stopping rule (e.g, $E_{\max}[\psi^{n+1}, \psi^n] < \varepsilon$, for some $\varepsilon > 0$) is met.

2.3 Edge-Featured Penalty and FFT

The previous dual approach for (2.9) requires to solve the nonlinear PDE (2.22), whose leading operator is highly nonlinear. In this section, we propose a new model for the dual variable with a linear leading operator, which allows for the use of more efficient numerical algorithms. The key idea is based on an edge-featured penalty of the inequality constraint $|\mathbf{p}| \leq 1$ in the definition of the TV-norm. It is important to observe that the dual variable $\mathbf{p}$ is actually an

Figure 2.2: Profiles of $|\nabla I|$, $|\nabla \psi|$ and $|\mathbf{p}|$ obtained by Algorithm 1 with $\lambda = 1, \mu = 1e3, r = 2, \tau = 0.1$ for $N = 100$ steps. Note: (i) the error $E_{\max}[\psi^N, \psi^{N-1}] = 2.1e - 7$, and (ii) as in Figure 2.1(f), ψ converges to (1.24), so we have approximately $|\nabla \psi| \leq 2\sqrt{2}$.

excellent edge detector. As shown in Figure 2.2, $|\mathbf{p}| \approx 1$ near the edge, while $|\mathbf{p}| \approx 0$ in the homogeneous region. Hence we can get (cf. [36])

$$|\mathbf{p}| = 1, \quad \text{if } \mathbf{H} \neq 0; \quad |\mathbf{p}| < 1, \quad \text{if } \mathbf{H} = 0. \quad (2.30)$$

Moreover, we find from (2.21), (2.23) and (2.29) that $\mathbf{H} = d \nabla \psi$. As ψ is designed to characterize the piecewise constant structures of the given image I , we have $|\mathbf{H}| = O(|\nabla \psi|) = O(|\nabla I|)$. Indeed, their profiles in Figure 2.2 justify this observation. Consequently, we treat the point-wise inequality constraint by penalizing $|\mathbf{p}| = 1$ strongly near sharp edges, and very weakly in the homogeneous regions. More precisely, we propose to approximate the subproblem (2.9) by the following minimization problem:

$$\min_{\psi} \max_{\mathbf{p}} \left\{ \int_{\Omega} \psi \operatorname{div} \mathbf{p} d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + \frac{\theta}{2} \int_{\Omega} (v - \psi)^2 d\mathbf{x} - \frac{1}{4} \int_{\Omega} g(|\nabla I|) (|\mathbf{p}|^2 - 1)^2 d\mathbf{x} \right\}, \quad (2.31)$$

where $g(s)$ is a smooth function of $s \in [0, \infty)$, such that $g(0) = 0$ and $g(s) \gg 1$ for $s \gg 1$. In this chapter, we take

$$g(s) = \alpha \ln(1 + s), \quad s \geq 0, \quad (2.32)$$

where α is a positive constant.

The optimality condition with respect to ψ yields $\psi = \mathcal{C}(\mathbf{p}, v, c_1, c_2)$ defined in (2.21), and the one with respect to $\mathbf{p}$ gives

$$-\nabla\psi + g(|\nabla I|)(1 - |\mathbf{p}|^2)\mathbf{p} = 0. \quad (2.33)$$

In view of (2.21) and (2.23), its gradient descent form is

$$\frac{\partial \mathbf{p}}{\partial t} = d^{-1} \mathbf{H}(\mathbf{p}, v, c_1, c_2) + g(|\nabla I|)(1 - |\mathbf{p}|^2)\mathbf{p}, \quad (2.34)$$

where $\mathbf{H}$ and d are given in (2.23) and (2.29), respectively. We can view (2.34) as a vectorial Allen-Cahn type model with $\nabla(\text{div})$ -diffusion.

To resolve (2.34) efficiently, we adopt a splitting scheme (or operator splitting, cf. [123, 83]), and solve the following two subproblems consecutively and recursively at each time step:

$$\frac{\partial \mathbf{p}}{\partial t} = d^{-1} \mathbf{H}(\mathbf{p}, v, c_1, c_2), \quad (2.35)$$

and

$$\frac{\partial \mathbf{p}}{\partial t} = g(|\nabla I|)(1 - |\mathbf{p}|^2)\mathbf{p}. \quad (2.36)$$

Let $\mathbf{p}^n$ be the approximation of $\mathbf{p}$ at $t_n = n\Delta t$. Formally, the second-order splitting scheme can be carried out as follows:

- (a). Solve the equation (2.36) for $t \in (t_n, t_n + \Delta t/2)$:

$$\tilde{\mathbf{p}} = \mathbf{p}^n \exp\left(g(|\nabla I|)(1 - |\mathbf{p}^n|^2)\frac{\Delta t}{2}\right). \quad (2.37)$$

- (b). Solve the equation (2.35) for $t \in (t_n, t_n + \Delta t)$ by using the Crank-Nicolson type scheme (see, e.g., [124]):

$$\frac{\hat{\mathbf{p}} - \tilde{\mathbf{p}}}{\Delta t} = \frac{1}{2d} \left(\mathbf{H}(\hat{\mathbf{p}}, v, c_1, c_2) + \mathbf{H}(\tilde{\mathbf{p}}, v, c_1, c_2) \right). \quad (2.38)$$

- (c). Find $\mathbf{p}^{n+1}$ by solving (2.36) for $t \in (t_n + \Delta t/2, t_{n+1})$:

$$\mathbf{p}^{n+1} = \hat{\mathbf{p}} \exp\left(g(|\nabla I|)(1 - |\hat{\mathbf{p}}|^2)\frac{\Delta t}{2}\right). \quad (2.39)$$

It is seen that with a linearization, the equation (2.36) becomes a simple ordinary differential equation. By imposing a periodic boundary condition for $\mathbf{p}$ (see [129, 138] for similar imposition), the linear equation (2.38) can be solved efficiently by using the Fourier-Galerkin method and FFT. One may refer to [118] for the implementation of FFT in Matlab.

We replace Step (iv) in Algorithm 1 by the splitting-FFT scheme and summarize the full algorithm for the minimization problem (2.8) in Algorithm 2 below.

Algorithm 2.

Just replace Step (iv) in **Algorithm 1** by the splitting-FFT scheme for (2.34), and remain the other steps. More precisely,

(iv) Compute $\mathbf{p}^{n+1}$ by

$$\begin{cases} \tilde{\mathbf{p}} = \exp\left(g(|\nabla I|)(1 - |\mathbf{p}^n|^2)\frac{\Delta t}{2}\right)\mathbf{p}^n, \\ -\frac{\Delta t}{2d^n}\nabla(\operatorname{div}\hat{\mathbf{p}}) + \hat{\mathbf{p}} = \tilde{\mathbf{p}} + \frac{\Delta t}{2d^n}\nabla(\operatorname{div}\tilde{\mathbf{p}}) - \frac{\Delta t}{d^n}s^n, \\ \mathbf{p}^{n+1} = \exp\left(g(|\nabla I|)(1 - |\hat{\mathbf{p}}|^2)\frac{\Delta t}{2}\right)\hat{\mathbf{p}}, \end{cases}$$

where

$$d^n = \frac{\lambda}{4}(c_1^n - c_2^n)^2 + \theta, \quad s^n = \frac{\lambda(c_1^n - c_2^n)}{2}\nabla I + \theta\nabla v^n.$$

The second equation is solved by the Fourier-Galerkin method with FFT.

We point out that due to the use of FFT, this algorithm is competitive to the primal approach and the dual Algorithm 1 in terms of computational complexity.

2.4 Numerical Results

The preliminary experiments in Section 2.1 have shown that the TV-Allen-Cahn type models (2.15) with various choices of r (i.e., “double-well” potentials) are

robust for binary image segmentation. In this section, we further test the models on several typical images with an emphasis on comparing the performance of the foregoing three algorithms, that is, the primal approach (2.19), the dual Algorithm 1 and the dual Algorithm 2. We assess their performance by examining the rate of convergence, decay of numerical energy and quality of segmentation.

In the first experiment, we still test the 100×100 annulus-triangle image with the same initial value (cf. Figure 2.1(d)), and take the parameters: $\lambda = 1, \mu = 500, r = 2$ in (2.15), $\alpha = 20$ in (2.32) and $\beta = 1e - 6$ in (2.19). To set up a relatively fair rule for comparison, we take the artificial time-marching step size τ or Δt to be comparable with values $1e - 3, 1e - 4, 5e - 5$ and $1e - 5$ for the primal approach and Algorithm 2. As the factor $d(\approx 2500)$ in (2.29) is scaled out in Algorithm 1 (cf. (2.24)), there holds the relation between the time steps of two algorithms: $\tau \approx d\Delta t$. Accordingly, we choose $\tau = 2.5, 0.25, 0.125$ and 0.025 in Algorithm 1.

We quantify the comparison by recording in Figure 2.3: (i) maximum pixel-wise errors between two consecutive steps: $E_{\max}\{\psi^{n+1}, \psi^n\}$; (ii) history of converging to the piecewise constant function (1.24) by examining $E\{|\psi^n|, 1\}$ (cf. (2.20)); and (iii) decay of numerical energy of the cost functional in (2.8).

We visualize from Figure 2.3 that two dual algorithms converge much faster than the primal approach. In other words, the phase function ψ evolves more rapidly to the targeted piecewise constant level-set function by dual methods. Moreover, we see that Algorithm 2 produces relatively accurate approximations than the other two methods. We remark that (i) for $\tau = 1e - 3$, the primal approach fails to converge due to the restriction of the CFL condition; (ii) the dual Algorithm 1 still converges with a choice of τ bigger than the critical value $1/8$ (cf. [34] for the ROF model). However, the convergence rate and accuracy deteriorate, when $\tau > 1/4$ (note: the choice of $\tau \leq 1/4$ is shown to be numerically stable in [34]).

Next, we test an image used in [22, 12] for discussing the global minimization

Figure 2.3: Comparison of the convergence rate with respect to the time-marching step n for three different time step sizes. Row 1: the maximum point-wise errors between two consecutive steps: $E_{\max}\{\psi^{n+1}, \psi^n\}$. Row 2: the convergence rate of $|\psi^n|$ to 1 measured by $E\{|\psi^n|, 1\}$. Row 3: Decay of numerical energy of the cost functional in (2.8) of three algorithms.

issue of the variational segmentation using CV-model and the classical level-set method. As shown in [12], the minimization algorithms based on some commonly-used explicit or implicit time evolution of the gradient descent flow often get stuck with a local minimizer, and therefore leads to a rather unsatisfactory segmentation of the narrow L -band in Figure 2.4(a). In this test, we choose the parameters and time steps in two dual algorithms as $\lambda = 1, \mu = 200, r = 2, \Delta t = 1e - 5, \tau = 1/8$ and $\alpha = 1e3$, and take two typical initial values (cf. Figure 2.4(a) and notice that the one at the bottom-left corner is similar to that in [22]).

Figure 2.4: Segmentation by Algorithms 1 and 2 with two different choices of initial values (note: the primal approach fails to converge for both options): (a) Given image I and two samples of initial values; (b) Segmented boundary Γ i.e., $\psi = 0$ at 130 steps; (c) profile of ψ at 130 steps.

We observe from Figure 2.4 that the numerical phase function ψ evolved by the dual algorithms converges very fast to the piecewise constant level-set function even with initial conditions far from the targets. Unfortunately, the primal approach (2.19) does not converge even for various attempts of different sets of parameters and time steps. However, the dual algorithms are very reliable and produce accurate segmentation with around 130 steps (see Figure 2.4(b)). As with the previous example, two dual algorithms enjoy almost the same convergence behavior, but Algorithm 2 seems to be slightly accurate. Hence, they are indistinguishable when we plot the boundaries of the segmented image, and the

profiles of ψ . In fact, similar performance is observed in the following tests involving relatively complicated images (see Figure 2.5 for the newspaper segmentation, and Figure 2.6 for the two-cell image with noise).

Figure 2.5: Segmentation by Algorithms 1 and 2 with $\lambda = 1, \mu = 1e3, r = 10, \alpha = 1e3, \tau = 1/8$ and $\Delta t = 1e - 5$: (a) Given image I and ψ^0 ; (b) Segmented boundary Γ i.e., $\psi = 0$ at 110 steps; (c) profile of ψ at 110 steps.

Figure 2.6: Segmentation by Algorithms 1 and 2 with the same parameters as in Figure 2.5: (a) Given image I and ψ^0 ; (b) Segmented image with marked Γ at 80 steps; (c) Profile of ψ at 80 steps.

In particular, we point out that the noisy two-cell image of the same size in Figure 2.6(a) was tested in [67], where the term $\int_{\Omega} (|\nabla \psi|^2 - 1)^2 d\mathbf{x}$ was added to penalize the classical level-set function so as to eliminate the re-initialization

process. The explicit evolution of the primal variable in [67] required at least 800 iterations to reach a meaningful segmentation. However, both dual algorithms provide accurate segmentation with 80 iterations, as shown in Figure 2.6.

(a) Given image I and initial value
 (b) u at 35 steps
 (c) u at 40 steps

Figure 2.7: Segmentation by Algorithms 2 with $\lambda = 1, \mu = 1e3, \tau = 1e - 5, \alpha = 2e3$ and $r = 10$: (a) Given image and two different initial conditions; (b) Segmented image u for the initial value given by bigger box at 35 steps; (c) Segmented image u for the initial value given by smaller box at 40 steps. Note: the errors $E\{|\psi^n|, 1\}$ for (b) and (c) are $1.16e - 1$ and $5.61e - 2$, respectively.

Finally, we use Algorithm 2 to segment the image in Figure 2.7(a) with quite a number of cells of different sizes but without noise. For two choices of initial values, the dual Algorithm 2 produces very fast and reliable segmentation. We see that though at around 40 steps, the errors $E\{|\psi^n|, 1\}$ are of accuracy about one or two significant digits, the segmentation of the set Γ is quite satisfactory.

2.5 Applications to Image Denoising

It is seen that Algorithm 2 is very efficient for the subproblem (2.9). It only requires to solve a vectorial Allen-Cahn type equation with leading linear differential operator, so more efficient algorithms can be implemented. The previous study and numerical results demonstrate that it provides a viable alternative to the Chambolle's algorithm [34]. We next show that it is also competitive for

solving the ROF model.

As the Chambolle's approach was originally proposed for the ROF model [112], it is desirable to compare the performance of both dual algorithms for image restoration. Given a noisy image I defined on an open bounded domain $\Omega \subseteq \mathbb{R}^2$, the ROF model restores I by solving the minimization problem:

$$\min_u \left\{ TV(u) + \frac{1}{2\delta} \int_{\Omega} |u - I|^2 d\mathbf{x} \right\}, \quad \delta > 0. \quad (2.40)$$

Like (2.9) and (2.31), a key idea to minimize (2.40) is to consider the following new minimization problem:

$$\min_u \max_{\mathbf{p}} \left\{ \int_{\Omega} u \operatorname{div} \mathbf{p} d\mathbf{x} + \frac{1}{2\delta} \int_{\Omega} |u - I|^2 d\mathbf{x} - \frac{1}{4} \int_{\Omega} g(|\nabla I|)(|\mathbf{p}|^2 - 1)^2 d\mathbf{x} \right\}. \quad (2.41)$$

Consequently, the optimality conditions lead to

$$u = I - \delta \operatorname{div} \mathbf{p}, \quad (2.42)$$

and

$$\nabla(\delta \operatorname{div} \mathbf{p} - I) + g(|\nabla I|)(1 - |\mathbf{p}|^2)\mathbf{p} = 0. \quad (2.43)$$

Like (2.34), we consider the gradient flow

$$\frac{\partial \mathbf{p}}{\partial t} = \delta \mathbf{H}(\mathbf{p}) + g(|\nabla I|)(1 - |\mathbf{p}|^2)\mathbf{p}, \quad (2.44)$$

where $\mathbf{H}(\mathbf{p}) = \nabla(\operatorname{div} \mathbf{p}) - \nabla I / \delta$. Similarly, it can be solved efficiently by using a time-splitting scheme and FFT as in (2.35)-(2.39).

Recall that in [34], the ROF model (2.40) is solved by the iteration:

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n + \tau \mathbf{H}(\mathbf{p}^n)}{1 + \tau |\mathbf{H}(\mathbf{p}^n)|}, \quad n = 0, 1, \dots, \quad (2.45)$$

and u is updated by (2.42).

We test two dual algorithms on lena-512 $\times$ 512, and as usual the peak signal-to-noise (PSNR) is used as a criteria for the quality of restoration:

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\frac{1}{mn} \sum_{i,j} (u_{i,j} - I_{i,j})^2}, \quad (2.46)$$

where $m \times n$ is the size of the image. Moreover, we use the relative dynamic error:

$$E_{l^2}(u^{n+1}, u^n) := \frac{\|u^{n+1} - u^n\|_{l^2}}{\|u^{n+1}\|_{l^2}} < \epsilon, \quad (2.47)$$

for a prescribed tolerance ϵ , as the stopping rule. We see from Figure 2.8 and the caption therein that with the same error tolerance, Algorithm 2 converges relatively faster with almost the same PSNR values. Such a behavior is consistent with the numerical experiments for image segmentation in the previous section.

(a) Original image (b) Noisy image I (c) u -Chambolle (d) u -new algorithm

Figure 2.8: Image restoration based on the ROF model with solutions by the Chambolle’s dual approach and the proposed new algorithm: (a) Original image: lena- 512×512 ; (b) Noisy image generated by the normal imnoise function in `Matlab` with “Gaussian”, zero mean and variance 0.02; (c)-(d): Restored images obtained by Chambolle’s method and the new dual algorithm with $\delta = 30, \tau = 1/8, \Delta t = 1e - 2$ and $g(s) = 0.6\sqrt[4]{s}$. The iteration is stopped by $E_{l^2}(u^{n+1}, u^n) \leq 1e - 3$. The iteration numbers and PSNR for (c)-Chambolle: 101 and 29.21, and for (d)-new algorithm: 89 and 29.16, respectively.

Global Minimization for Multiphase Labeling and Segmentation Problems

The multiclass labeling and multiphase segmentation problems share some similarity in nature, as typically both of them aim to find a partition of an image into m disjoint regions (phases or classes) according to some optimization rule. As a matter of fact, the former term is commonly used in computer vision, while the latter one often appears in the variational and PDE community.

Accordingly, many recent attempts tackle this problem from a continuous point of view by minimizing the continuous Potts model (1.13) (see, e.g., [64, 107, 104, 24, 14] and the references therein). An important issue on solving (1.13) is to find a convex approximation of the non-convex problem:

$$\min_{\{v_i\}_{i=1}^m} \left\{ E(\{v_i\}_{i=1}^m) := \lambda \sum_{i=1}^m \int_{\Omega} f_i(\mathbf{x}) v_i(\mathbf{x}) d\mathbf{x} + \sum_{i=1}^m TV(v_i) \right\}, \quad (3.1)$$

subject to $v_i \in \{0, 1\}$ and $\sum_{i=1}^m v_i = 1$ on Ω , where $\{v_i\}$ are the labeling functions and the total variation (TV) is defined in Chapter 1. The optimum v_i is expected to be the indicator function χ_i of Ω_i , i.e., a binary value.

Indeed, a major class of methods is based on the convex relaxation of the admissible set by allowing for the labeling functions to take “intermediate” values

from the unit simplex:

$$\Delta_+ := \left\{ \mathbf{v} := (v_1, \dots, v_m) \in \mathbb{R}^m : v_i \in [0, 1] \text{ and } \sum_{i=1}^m v_i = 1, \forall \mathbf{x} \in \Omega \right\}. \quad (3.2)$$

Here, we just mention a few contributions along this line. Zack et al. [144] minimized the energy in (3.1) over (3.2) by introducing additional variables to decouple the TV-term and the data term, while Lellmann et al. [64] replaced the TV-term by non-isotropic TV: $\int_{\Omega} (\sum_i |\nabla v_i|^2)^{1/2} d\mathbf{x}$, and applied a Douglas-Rachford splitting algorithm for solving the modified model (also refer to [65] for an up-to-date review). Pock et al. [107] proposed a primal-dual projected gradient algorithm for minimizing an energy functional different from (3.1) that produced a tighter bound on the energy and often (computationally) led to the global minimizer. Li et al. [70, 68] applied the same labeling technique to the piecewise constant Mumford-Shah-type model with data fidelity involving fuzzy membership functions, and minimized the resulting model by using alternative minimization methods as in [22]. More recently, Bae, Yuan and Tai [14] proposed a dual-type method for the approximate model:

$$\min_{\mathbf{v} \in \Delta_+} E_{\varepsilon}(\mathbf{v}) := \sum_{i=1}^m \left\{ TV(v_i) + \int_{\Omega} (\lambda f_i v_i + \varepsilon v_i \log v_i) d\mathbf{x} \right\}, \quad (3.3)$$

where $\varepsilon > 0$ is sufficiently small. The additional log-sum term can be essentially viewed as a penalization of the constraint $v_i = 0, 1$. Sufficient conditions for the existence of global optimum solution of the primal-dual model were derived. Moreover, the comprehensive comparison conducted in [14] demonstrated the dual-type approach outperformed most of the existing approaches mentioned above, and could achieve global minimizers at least computationally. In [75], the TV-term in (3.1) was replaced by another term which approximates the length term. Afterward, a special linearization technique was used to derive an algorithm that was shown to be fast without solving any equations (since all subproblems admitted explicit solutions, and certain narrow-band technique could be applied to accelerate the algorithm).

Another family of methods based on convex formulation via functional lifting and embedding in a higher-dimensional space has also attracted much recent attention, see, e.g., [96, 107, 24, 25, 106].

This chapter is motivated by [14], and the main features and contributions can be summarized as follows.

- Different from most of the previous works, we stick with the binary setting (i.e., without relaxing the admissible set as in (3.2)) by seeking the minimizer from

$$\mathcal{A} := \left\{ \mathbf{v} = (v_1, v_2, \dots, v_m) : v_i^2 = v_i, \sum_{i=1}^m v_i = 1 \right\}, \quad (3.4)$$

and conduct a direct analysis for the primal-dual formulation of (3.1):

$$\min_{\mathbf{v} \in \mathcal{A}} \max_{\vec{\mathbf{p}} \in \mathbf{S}} \left\{ E(\mathbf{v}, \vec{\mathbf{p}}) := \sum_{i=1}^m \int_{\Omega} v_i (\operatorname{div} \mathbf{p}_i + \lambda f_i) d\mathbf{x} \right\}, \quad (3.5)$$

where $\vec{\mathbf{p}} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m)$ and $\mathbf{S} = S^m$ (see (1.5) for the definition of S). We provide the sufficient and necessary conditions for finding the global optimum of the primal-dual model (3.5). Although some results have been derived earlier in [14], the arguments and means are quite different.

- Based on the augmented Lagrangian method, we obtain an explicit expression of the primal variables in terms of the dual variables, where the primal variables take the binary values in most of the situations. As a result, the thresholding is not necessary in this case. Moreover, the iterative algorithm contains almost the minimum number of parameters and is easy to initialize. The essential step is very analogous to the Chambolle's algorithm [34] for image denoising, so the method is expected to be fast and robust.
- The approach can also be interpreted as a multiphase level-set method based on piecewise constant interpolation of the phases (or classes), so it is anticipated to be stabler than the classical level-set method [135] and the piecewise constant level-set method [73] (a global polynomial interpolation).

The rest of the chapter is organized as follows. We formulate and analyze the algorithm in the forthcoming section. More precisely, we derive the sufficient and necessary conditions for finding global optimum of the primal-dual model (3.5) without using convex relaxation. We develop the algorithm which can oftentimes achieve the global optimum. We conduct a comparison study and provide ample numerical results to show the strengths and performance of the proposed method for multiclass labeling and multiphase image segmentation in Section 3.2.

3.1 Analysis of the Primal-Dual Model and the Algorithm

In this section, we conduct analysis of the global optimum of the primal-dual model (3.5), and introduce the fast algorithms. Due to similar nature of the labeling and segmentation problems, we shall not distinguish the words “class” and “phase”, and likewise “multiclass” and “multiphase” throughout the chapter.

3.1.1 Two-phase case

To provide some insights of the algorithms for multiphase problems, we first discuss the two-phase case. In this situation, the admissible set (3.4) can be characterized by a single labeling function v_1 as $v_2 = 1 - v_1$. For convenience, we define $v = 2v_1 - 1$, so we have $v_1 = (1 + v)/2$ and $v_2 = (1 - v)/2$. Accordingly, the constraint on v_1 and v_2 becomes $v^2 = 1$, and by (3.1),

$$\begin{aligned} & \min_{v^2=1} \left\{ \lambda \int_{\Omega} \left(f_1 \frac{1+v}{2} + f_2 \frac{1-v}{2} \right) d\mathbf{x} + TV(v) \right\} \\ &= \min_{v^2=1} \left\{ \lambda \int_{\Omega} g v d\mathbf{x} + TV(v) + \lambda \int_{\Omega} \frac{f_1 + f_2}{2} d\mathbf{x} \right\}, \end{aligned} \quad (3.6)$$

where $g = (f_1 - f_2)/2$. Thus, by dropping the last term, we obtain from the definition of total variation in (1.26) that

$$\min_{v^2=1} \max_{\mathbf{p} \in S} \{ L(v, \mathbf{p}) := \langle v, \operatorname{div} \mathbf{p} + \lambda g \rangle \}, \quad (3.7)$$

where $\langle u, v \rangle = \int_{\Omega} uv \, d\mathbf{x}$ and S is the same as in (1.5).

The global optimal v is the binary value ± 1 . The following characterization is essential for the development of the algorithm.

Theorem 3.1. *Let*

$$\mathbf{p}^* = \arg \min_{\mathbf{p} \in S} \int_{\Omega} |\operatorname{div} \mathbf{p} + \lambda g| \, d\mathbf{x}. \quad (3.8)$$

Then the pair $(v^, \mathbf{p}^*)$ is the optimum of the primal-dual problem (3.7), if and only if $(v^*, \mathbf{p}^*)$ satisfies*

$$|v^*| = 1, \quad (\operatorname{div} \mathbf{p}^* + \lambda g) + |\operatorname{div} \mathbf{p}^* + \lambda g| v^* = 0 \text{ a.e. on } \Omega. \quad (3.9)$$

Proof. Notice that

$$L(v, \mathbf{p}) \geq -\|v\|_{L^\infty(\Omega)} \|\operatorname{div} \mathbf{p} + \lambda g\|_{L^1(\Omega)},$$

where $\|\cdot\|_{L^p(\Omega)}$ is the usual L^p -norm. Therefore, for any v in the admissible set, we have

$$\max_{\mathbf{p} \in S} L(v, \mathbf{p}) \geq \max_{\mathbf{p} \in S} (-\|\operatorname{div} \mathbf{p} + \lambda g\|_{L^1(\Omega)}) = -\min_{\mathbf{p} \in S} \|\operatorname{div} \mathbf{p} + \lambda g\|_{L^1(\Omega)}.$$

Therefore, the optimum is attained if and only if the equality holds, that is, the second identity of (3.9) is true. $\square$

Remark 3.1. *It is worthwhile to point out that the solution of (3.7) is not unique. Indeed, the primal-dual problem (3.7) is equivalent to*

$$\min_{v^2=1} \max_{\mathbf{p} \in \tilde{S}} \{L(v, \mathbf{p}) := \langle v, \operatorname{div} \mathbf{p} + \lambda g \rangle\}, \quad (3.10)$$

where $\tilde{S}$ is a subset of S in (1.5) defined by

$$\tilde{S} := \left\{ \mathbf{p} = (p_1, p_2) \in C_c^1(\Omega; \mathbb{R}^2) : |\mathbf{p}| \leq 1 \text{ and } \operatorname{div} \mathbf{p} + \lambda g \neq 0, \forall \mathbf{x} \in \Omega \right\}.$$

Therefore, v^* is expressed by $\mathbf{p}^* \in \tilde{S}$ via

$$v^* = -\frac{\operatorname{div} \mathbf{p}^* + \lambda g}{|\operatorname{div} \mathbf{p}^* + \lambda g|}, \quad (3.11)$$

which indicates that v^* takes the sign of $-(\operatorname{div} \mathbf{p}^* + \lambda g)$. $\square$

Remark 3.2. *The situation is reminiscent to the Chambolle's dual algorithm [34] for the Rudin-Osher-Fatemi (ROF) model [112]:*

$$\min_u \left\{ TV(u) + \frac{\mu}{2} \|u - I\|_{L^2(\Omega)}^2 \right\}, \quad \mu > 0, \quad (3.12)$$

where I is a given noisy image. In this case, the primal-dual problem takes the form

$$\min_u \max_{\mathbf{p} \in S} \int_{\Omega} \left(u \operatorname{div} \mathbf{p} + \frac{\mu}{2} |u - I|^2 \right) d\mathbf{x}. \quad (3.13)$$

The dual algorithm in [34] is essentially based on minimizing the dual problem:

$$\min_{\mathbf{p} \in S} \int_{\Omega} |\operatorname{div} \mathbf{p} - \mu I|^2 d\mathbf{x}, \quad (3.14)$$

by solving the nonlinear equation:

$$-\nabla(\operatorname{div} \mathbf{p} - \mu I) + |\nabla(\operatorname{div} \mathbf{p} - \mu I)| \mathbf{p} = 0. \quad (3.15)$$

For clarity, we sketch the derivation of (3.15) from (3.14). Following [34] (also see [41]), we take the Lagrangian $\hat{L}[\mathbf{p}, \alpha]$ of the dual problem (3.14) with the Lagrange multiplier $\alpha = \alpha(\mathbf{x})$:

$$\hat{L}[\mathbf{p}, \alpha] = \int_{\Omega} \left(\mu I \operatorname{div} \mathbf{p} - \frac{1}{2} (\operatorname{div} \mathbf{p})^2 + \frac{\alpha}{2} (1 - |\mathbf{p}|^2) \right) d\mathbf{x}. \quad (3.16)$$

By the optimality condition with respect to $\mathbf{p}$,

$$-\nabla(\operatorname{div} \mathbf{p} - \mu I) + \alpha \mathbf{p} = 0. \quad (3.17)$$

The complementary condition of the Lagrange multiplier implies that if $|\mathbf{p}| = 1$ at the optimum, then $\alpha > 0$, whereas if $|\mathbf{p}| < 1$, then $\alpha = 0$. In any case, we have $\alpha = |\nabla(\operatorname{div} \mathbf{p} - \mu I)|$. This gives (3.15).

It is important to notice the difference between (3.8) and (3.14), that is, L^1 -minimization (due to the presence of the constraint $v^2 = 1$ in (3.7)) versus L^2 -minimization.

In [136], a different principle (cf. [22]) was adopted to derive a Chambolle-type algorithm for two-phase segmentation. However, the current approach appears more natural. $\square$

Now, we follow the procedure as above to solve the dual problem (3.8). Like (3.16), we define

$$\tilde{L}[\mathbf{p}, \gamma] = \int_{\Omega} \left(|\operatorname{div} \mathbf{p} + \lambda g| + \frac{\gamma}{2} (|\mathbf{p}|^2 - 1) \right) d\mathbf{x}.$$

This leads to the counterpart of (3.17):

$$-\nabla \left(\frac{\operatorname{div} \mathbf{p} + \lambda g}{|\operatorname{div} \mathbf{p} + \lambda g|} \right) + \gamma \mathbf{p} = 0. \quad (3.18)$$

By the same argument (see the description below (3.17)), we obtain $\gamma = \left| \nabla \left(\frac{\operatorname{div} \mathbf{p} + \lambda g}{|\operatorname{div} \mathbf{p} + \lambda g|} \right) \right|$. Inserting it into (3.18) yields the nonlinear equation analogous to (3.15):

$$-\nabla \left(\frac{\operatorname{div} \mathbf{p} + \lambda g}{|\operatorname{div} \mathbf{p} + \lambda g|} \right) + \left| \nabla \left(\frac{\operatorname{div} \mathbf{p} + \lambda g}{|\operatorname{div} \mathbf{p} + \lambda g|} \right) \right| \mathbf{p} = 0, \quad (3.19)$$

which can be solved by a gradient descent approach. More precisely, we consider the gradient decent flow of (3.19):

$$\frac{\partial \mathbf{p}}{\partial t} = -\nabla v - |\nabla v| \mathbf{p} \quad \text{with} \quad v = -\frac{\operatorname{div} \mathbf{p} + \lambda g}{|\operatorname{div} \mathbf{p} + \lambda g|}, \quad (3.20)$$

where the last identity is due to (3.9), and with a little abuse of notation, we still use $\mathbf{p}$ and v to denote the unknowns though they depend on the artificial variable t . Let τ be the time step size and $\mathbf{p}^n$ be the approximation of $\mathbf{p}$ at $t = n\tau$. To avoid division by zero, we adopt the conventional regularization (cf. [97]) and define

$$v^n := -\frac{\operatorname{div} \mathbf{p}^n + \lambda g}{|\operatorname{div} \mathbf{p}^n + \lambda g|_{\beta}}, \quad (3.21)$$

where $|\operatorname{div} \mathbf{p}^n + \lambda g|_{\beta} = |\operatorname{div} \mathbf{p}^n + \lambda g| + \beta$ for some sufficiently small $\beta > 0$. Then, we resort to the semi-implicit discretization in time as in [34], to solve (3.20):

$$\frac{\mathbf{p}^{n+1} - \mathbf{p}^n}{\tau} = -\nabla v^n - |\nabla v^n| \mathbf{p}^{n+1} \Rightarrow \mathbf{p}^{n+1} = \frac{\mathbf{p}^n - \tau \nabla v^n}{1 + \tau |\nabla v^n|}. \quad (3.22)$$

It is seen that the L^1 -minimization (3.8) induces additional nonlinearity, compared with the L^2 -minimization (3.13).

In view of Remark 3.1, we adopt the Merriman-Bence-Osher (MBO)-type projection (see, e.g. [86, 128])

$$\mathcal{P}_B(\kappa) := \begin{cases} 1, & \text{if } \kappa \geq 0, \\ -1, & \text{if } \kappa < 0, \end{cases} \quad (3.23)$$

to obtain the binary value v .

We summarize the algorithm as follows, where the involved differential operators can be discretized as in [34].

Algorithm 1

1. Initialization: set $\mathbf{p}^0 = \mathbf{0}$ and choose $\beta, \tau > 0$;

2. For $n = 0, 1, \dots$

(i) Compute

$$v^n := -\frac{\operatorname{div} \mathbf{p}^n + \lambda g}{|\operatorname{div} \mathbf{p}^n + \lambda g|_\beta};$$

(ii) Update $\mathbf{p}$ by the Chambolle-type algorithm:

$$\mathbf{p}^{n+1} = \frac{\mathbf{p}^n - \tau \nabla v^n}{1 + \tau |\nabla v^n|};$$

3. Endfor till some stopping rule meets;

4. Set

$$v = \mathcal{P}_B(v^n).$$

3.1.2 Multiphase case

With the insights from two-phase model, we now consider the full primal-dual model (3.5). Observe that for any $\mathbf{v} \in \mathcal{A}$, we have

$$\begin{aligned} \sum_{i=1}^m v_i (\operatorname{div} \mathbf{p}_i + \lambda f_i) &\geq \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i + \lambda f_i \} \sum_{i=1}^m v_i \\ &= \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i + \lambda f_i \}, \quad \forall \mathbf{x} \in \Omega, \end{aligned} \quad (3.24)$$

which implies

$$E(\mathbf{v}, \vec{\mathbf{p}}) \geq \int_{\Omega} \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i + \lambda f_i \} d\mathbf{x}, \quad (3.25)$$

for all $(\mathbf{v}, \vec{\mathbf{p}}) \in \mathcal{A} \times \mathcal{S}$, where the energy functional $E(\cdot, \cdot)$ is given in (3.5). Hence, we deduce from (3.25) that

$$\max_{\vec{\mathbf{p}} \in \mathcal{S}} E(\mathbf{v}, \vec{\mathbf{p}}) \geq \max_{\vec{\mathbf{p}} \in \mathcal{S}} \left\{ E_D(\vec{\mathbf{p}}) := \int_{\Omega} \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i + \lambda f_i \} d\mathbf{x} \right\}, \quad \forall \mathbf{v} \in \mathcal{A}. \quad (3.26)$$

Accordingly, the global optimum $(\mathbf{v}^*, \vec{\mathbf{p}}^*)$ of the primal-dual model (3.5) is a pair in $\mathcal{A} \times \mathcal{S}$ such that the identity of (3.25) holds, and the global minimizer $\mathbf{v}^*$ of the original model (3.1) is in $\mathcal{A}$ such that the identity (3.26) holds.

Similar to Theorem 1 in [14], we have the following result on the characterization of the global optimum $(\mathbf{v}^*, \vec{\mathbf{p}}^*)$.

Theorem 3.2. *Let*

$$\vec{\mathbf{p}}^* = \arg \max_{\vec{\mathbf{p}} \in \mathcal{S}} E_D(\vec{\mathbf{p}}). \quad (3.27)$$

- Suppose that $\min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \}$ has a unique minimum value for all $\mathbf{x} \in \Omega$. Then $\mathbf{v}^*$ must take the form

$$v_k^* = \begin{cases} 1, & k = \arg \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \}, \\ 0, & \text{otherwise,} \end{cases} \quad (3.28)$$

for $1 \leq k \leq m$.

- Suppose that $\min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \}$ has more than one minimum values at some $\mathbf{x} \in \Omega$, say, k values: $\operatorname{div} \mathbf{p}_j^* + \lambda f_j, j \in \{j_1, j_2, \dots, j_k\}$. Then $\mathbf{v}^*$ must satisfy

$$\sum_{i=1}^k v_{j_i}^* = 1 \quad \text{and} \quad v_j^* = 0, \quad j \notin \{j_1, \dots, j_k\}. \quad (3.29)$$

Proof. Since the first statement is a special case of the second one, it suffices to prove the second statement. By (3.25), we have $E(\mathbf{v}, \vec{\mathbf{p}}^*) \geq E_D(\vec{\mathbf{p}}^*)$ for any

$\mathbf{v} \in \mathcal{A}$. Thus it is enough to show that $E(\mathbf{v}^*, \vec{\mathbf{p}}^*) = E_D(\vec{\mathbf{p}}^*)$. It is clear that if $\mathbf{v}^*$ is defined by (3.29), then

$$\begin{aligned} \sum_{i=1}^m v_i^* (\operatorname{div} \mathbf{p}_i^* + \lambda f_i) &= \sum_{i=1}^k v_{j_i}^* (\operatorname{div} \mathbf{p}_{j_i}^* + \lambda f_{j_i}) \\ &= \sum_{i=1}^k v_{j_i}^* \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \} = \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \}, \end{aligned}$$

for $\mathbf{x} \in \Omega$ and the corresponding $1 \leq k \leq m$. Integrating the above identities over Ω leads to $E(\mathbf{v}^*, \vec{\mathbf{p}}^*) = E_D(\vec{\mathbf{p}}^*)$.

Next, we prove that (3.29) is necessary. If there is a v_l^* with $l \in \{j_1, j_2, \dots, j_k\}$, which does not correspond to a minimum value, i.e., $\operatorname{div} \mathbf{p}_l^* + \lambda f_l > \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \}$, then

$$\begin{aligned} \sum_{i=1}^m v_i^* (\operatorname{div} \mathbf{p}_i^* + \lambda f_i) &= \sum_{i=1}^k v_{j_i}^* (\operatorname{div} \mathbf{p}_{j_i}^* + \lambda f_{j_i}) \\ &= (1 - v_l^*) \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \} + v_l^* (\operatorname{div} \mathbf{p}_l^* + \lambda f_l) \\ &= \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \} + v_l^* (\operatorname{div} \mathbf{p}_l^* + \lambda f_l - \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \}). \end{aligned}$$

Therefore, the optimum value can be obtained if and only if $v_l^* = 0$. $\square$

Remark 3.3. *It is worthwhile to point out that the necessity of (3.28) and (3.29) is under the distributional sense. Indeed, we are free to change the value of $\mathbf{v}^*$ at a measure zero set of Ω without affecting the integral value.* $\square$

Remark 3.4. *A similar analysis was conducted in [14] (see Theorem 1 of [14]) based on convex relaxation. Here, we provided a direct concise proof, which did not require to switch the min-max using Karush-Kuhn-Tucher (KKT) conditions [60] and thresholding techniques. On the other hand, we could show that (3.29) is also necessary.* $\square$

We find from (3.27) that it is essential to solve the dual problem defined in (3.24). However, it is nonsmooth, so it appears very challenging to work directly

on this dual model. The following smoothed-dual model was used in [14]:

$$\max_{\vec{p} \in \mathcal{S}} \left\{ E_{D,\varepsilon}(\vec{p}) := -\varepsilon \int_{\Omega} \log \sum_{i=1}^m \exp \left(\frac{-\operatorname{div} \mathbf{p}_i - \lambda f_i}{\varepsilon} \right) d\mathbf{x} \right\}, \quad (3.30)$$

where $0 < \varepsilon \ll 1$. It is important to note that (3.30) turns out to be the dual model of (3.3).

Hereafter, we shall take a different approach and derive the algorithm based on the augmented Lagrangian method (ALM). It is worthwhile to point out the ALM has been recently widely used for solving minimization problems in image processing (see, e.g., [138, 129, 142] and the references therein). Here, we use it as a tool to derive the relation between the primal and dual variables via eliminating all the intermediate parameters and unknowns. This leads to an algorithm that involves the minimum number of parameters, and can be implemented as efficiently as the Chambolle's algorithm in [34].

3.1.3 Description of the algorithm

We start with (3.1) and rewrite it as

$$\min_v \left\{ \sum_{i=1}^m \int_{\Omega} |\nabla v_i| d\mathbf{x} + \lambda \sum_{i=1}^m \int_{\Omega} f_i v_i d\mathbf{x} \right\}.$$

As with the ALM for the ROF model (see, e.g., [138, 129]), an essential step for dealing with the TV-term is to introduce an intermediate variable and reformulate the above problem into the following constrained problem:

$$\begin{aligned} \min_{\vec{q}, v} & \left\{ \sum_{i=1}^m \int_{\Omega} |\mathbf{q}_i| d\mathbf{x} + \lambda \sum_{i=1}^m \langle f_i, v_i \rangle \right\}, \\ \text{subject to } & \mathbf{q}_i = \nabla v_i, v_i^2 = v_i, \sum_{i=1}^m v_i = 1. \end{aligned}$$

Note that the gradient operator on functions of bounded variation (BV) should be understood in the weak sense (cf. [52]). Using the notion of ALM yields the

unconstrained problem:

$$\begin{aligned}
\min_{\mathbf{v}, \vec{\mathbf{q}}} \max_{\vec{\boldsymbol{\lambda}}_1, \boldsymbol{\lambda}_2, \lambda_3} \Big\{ & \mathcal{L}(\mathbf{v}, \vec{\mathbf{q}}, \vec{\boldsymbol{\lambda}}_1, \boldsymbol{\lambda}_2, \lambda_3) := \sum_{i=1}^m \left(\langle 1, |\mathbf{q}_i| \rangle + \lambda \langle f_i, v_i \rangle + \langle \boldsymbol{\lambda}_{1i}, \mathbf{q}_i - \nabla v_i \rangle \right. \\
& + \frac{r_1}{2} \langle \mathbf{q}_i - \nabla v_i, \mathbf{q}_i - \nabla v_i \rangle + \langle \lambda_{2i}, v_i^2 - v_i \rangle + \frac{r_2}{2} \langle v_i^2 - v_i, v_i^2 - v_i \rangle \Big) \\
& \left. + \left\langle \lambda_3, \sum_{i=1}^m v_i - 1 \right\rangle + \frac{r_3}{2} \left\langle \sum_{i=1}^m v_i - 1, \sum_{i=1}^m v_i - 1 \right\rangle \right\},
\end{aligned} \tag{3.31}$$

where $r_1, r_2, r_3 > 0$ are penalization constants, and the vector-valued functions $\vec{\boldsymbol{\lambda}}_1 = (\boldsymbol{\lambda}_{11}, \dots, \boldsymbol{\lambda}_{1m})$, $\boldsymbol{\lambda}_2 = (\lambda_{21}, \dots, \lambda_{2m})$ (with $\lambda_{2i} \geq 0$) and the scalar function λ_3 are Lagrange multipliers. The optimality conditions lead to the system:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial v_i} &= \lambda f_i + \operatorname{div} \boldsymbol{\lambda}_{1i} + r_1 \operatorname{div}(\mathbf{q}_i - \nabla v_i) + \lambda_{2i}(2v_i - 1) \\
&+ r_2(v_i^2 - v_i)(2v_i - 1) + \lambda_3 + r_3 \left(\sum_{i=1}^m v_i - 1 \right) = 0,
\end{aligned} \tag{3.32a}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} = \frac{\mathbf{q}_i}{|\mathbf{q}_i|} + \boldsymbol{\lambda}_{1i} + r_1(\mathbf{q}_i - \nabla v_i) = 0, \tag{3.32b}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}_{1i}} = \mathbf{q}_i - \nabla v_i = 0, \tag{3.32c}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_{2i}} = v_i^2 - v_i = 0, \tag{3.32d}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_3} = \sum_{i=1}^m v_i - 1 = 0. \tag{3.32e}$$

Thus, it follows from (3.32b) and (3.32c) that

$$\boldsymbol{\lambda}_{1i} = -\frac{\nabla v_i}{|\nabla v_i|}, \quad 1 \leq i \leq m. \tag{3.33}$$

In fact, $\vec{\boldsymbol{\lambda}}_1$ turns out to be the vector of dual variables. To develop an efficient algorithm, it is essential to express the primal variable $\mathbf{v}$ in terms of $\vec{\boldsymbol{\lambda}}_1$. Using (3.32c), (3.32d) and (3.32e), we can simplify (3.32a) into

$$\lambda_{2i}(2v_i - 1) = -(\lambda f_i + \operatorname{div} \boldsymbol{\lambda}_{1i} + \lambda_3), \tag{3.34}$$

which implies

$$\lambda_{2i}^2(4v_i^2 - 4v_i + 1) = (\lambda f_i + \operatorname{div} \boldsymbol{\lambda}_{1i} + \lambda_3)^2.$$

Then by (3.32d) (note: $\lambda_{2i} \geq 0$),

$$\lambda_{2i} = |\lambda f_i + \operatorname{div} \boldsymbol{\lambda}_{1i} + \lambda_3| = |h_i + \lambda_3|,$$

where $h_i := \operatorname{div} \boldsymbol{\lambda}_{1i} + \lambda f_i$. Plugging it into (3.34) leads to

$$(2v_i - 1)|h_i + \lambda_3| + (h_i + \lambda_3) = 0. \quad (3.35)$$

It is seen that the use of the ALM allows us to derive the (reduced) primal-dual system (3.33) and (3.35) with a free variable λ_3 . In contrast with the usual Uzawa algorithm for (3.31), our algorithm based on (3.33) and (3.35) involves the minimum number of parameters and variables, so the iterative solver is easy to initialize, as to be shown shortly.

An essential step is to choose λ_3 and express the primal variables $\{v_i\}$ in terms of $\{\boldsymbol{\lambda}_{1i}\}$. For this purpose, let h_k (resp. h_j) be (one of) the smallest (resp. the second smallest) value of $\{h_i\}_{i=1}^m$ (note: if all $\{h_i\}_{i=1}^m$ are equal, then $h_j = h_k$). By choosing

$$-\lambda_3 = h_k + \frac{h_j - h_k}{2}, \quad (3.36)$$

we find from (3.35) that $v_k = 1$ and $v_i = 0$ for all the i s such that $h_i \leq h_j$. It is worthwhile to point out that (3.35) can not identify v_i when $h_i + \lambda_3 = 0$, i.e., $h_j = h_k$. In the computation, we regularize (3.35) and find

$$v_i = \frac{1}{2} - \frac{h_i + \lambda_3}{2|h_i + \lambda_3|_\beta}, \quad (3.37)$$

where λ_3 is defined in (3.36), and $|h_i + \lambda_3|_\beta = |\operatorname{div} \boldsymbol{\lambda}_{1i} + \lambda f_i + \lambda_3| + \beta$ for sufficiently small $\beta > 0$ as in (3.21).

Next, for fixed $\mathbf{v}$, the equation (3.33) can be solved by applying the iterative algorithm in [34] to

$$\nabla v_i + |\nabla v_i| \boldsymbol{\lambda}_{1i} = \mathbf{0}, \quad 1 \leq i \leq m, \quad (3.38)$$

with $\vec{\lambda}_1 \in \mathbf{S}$, as described in (3.19)-(3.22). Note that $\vec{\lambda}_1$ is the “approximation” of $\vec{p}$ in Subsection 3.1.2.

Remark 3.5. *As predicted by Theorem 3.2, if $\vec{\lambda}_1^*$ solves (3.27), then by defining $\mathbf{v}^*$ as*

$$v_k^* = \begin{cases} 1, & k = \min \{ \arg \min_{1 \leq i \leq m} \{ \operatorname{div} \lambda_{1i}^* + \lambda f_i \} \}, \\ 0, & \text{otherwise,} \end{cases}$$

for all $1 \leq k \leq m$ and $\mathbf{x} \in \Omega$, we have the following properties.

- (i) $\mathbf{v}^*$ satisfies the optimality equation (3.35) with $\lambda_3 = -(\operatorname{div} \lambda_{1k}^* + \lambda f_k)$.
- (ii) Under the assumption of (3.28), that is, $\min_{1 \leq i \leq m} \{ \operatorname{div} \lambda_{1i}^* + \lambda f_i \}$ has a unique minimum value for all $\mathbf{x} \in \Omega$, we find from Theorem 3.2 that $\mathbf{v}^*$ is expected to be the unique global minimizer.
- (iii) Only one component of $\mathbf{v}^*$ takes the value 1, while the other $m - 1$ components are all zero. According to Theorem 3.2, the so-defined $\mathbf{v}^*$ is expected to be a global minimizer, since $\mathbf{v}^*$ satisfies (3.29). $\square$

Now, we are ready to present the full algorithm as follows.

Algorithm 2

1. Initialization: set $\vec{\lambda}_1^0 = \mathbf{0}$ and choose $\beta, \tau > 0$;
2. For $n = 0, 1, \dots$
 - (i) Compute λ_3 by (3.36);
 - (ii) Compute $\mathbf{v}^n$ by (3.37);
 - (iii) Compute $\vec{\lambda}_1^{n+1}$ by the Chambolle-type algorithm:

$$\lambda_{1i}^{n+1} = \frac{\lambda_{1i}^n - \tau \nabla v_i^n}{1 + \tau |\nabla v_i^n|}, \quad 1 \leq i \leq m;$$

3. Endfor till some stopping rule meets;

4. Set $\vec{\lambda}_1^* = \vec{\lambda}_1^{n+1}$ and define

$$v_k^* = \begin{cases} 1, & k = \min \{ \arg \min_{1 \leq i \leq m} \{ \operatorname{div} \lambda_{1i}^* + \lambda f_i \} \}, \\ 0, & \text{otherwise,} \end{cases}$$

for all $1 \leq k \leq m$.

3.2 A Comparison Study and Numerical Results

In this section, we demonstrate the performance of the proposed algorithm by testing it on some typical images and comparing it with some relevant methods.

3.2.1 A comparison study

To provide more insights, we compare the algorithm with the global smoothed-dual algorithm in [14], and examine the approach from the perspective of the level set methodology.

1. Global smoothed-dual algorithm

The dual approach in [14] is essentially based on the model (3.3), which approximates the original non-convex problem (3.1) when $0 < \varepsilon \ll 1$. The optimum of the dual variable is obtained by solving the model (3.30), which is a smoothed version of the non-smooth dual problem (3.27). This leads to the representation of the primal variables $\{v_i\}$ in terms of the dual variables $\{\mathbf{p}_i\}$:

$$v_i^\varepsilon = \frac{\exp\left(-\frac{\operatorname{div} \mathbf{p}_i + \lambda f_i}{\varepsilon}\right)}{\sum_{j=1}^m \exp\left(-\frac{\operatorname{div} \mathbf{p}_j + \lambda f_j}{\varepsilon}\right)} = \frac{1}{\sum_{j=1}^m \exp\left(-\frac{(\operatorname{div} \mathbf{p}_j + \lambda f_j) - (\operatorname{div} \mathbf{p}_i + \lambda f_i)}{\varepsilon}\right)}, \quad (3.39)$$

for $1 \leq i \leq m$ and $\mathbf{x} \in \Omega$. Apparently, the vector $\mathbf{v}^\varepsilon$ with components $\{v_i^\varepsilon\}$ is in the convex simplex Δ_+ in (3.2). Suppose that $\operatorname{div} \mathbf{p}_i + \lambda f_i$ is the unique minimum

value of $\{\operatorname{div} \mathbf{p}_j + \lambda f_j\}_{j=1}^m$. Then we observe from (3.39) that $v_i^\varepsilon \rightarrow 1$ and $v_j^\varepsilon \rightarrow 0$ (for all $j \neq i$) as $\varepsilon \rightarrow 0^+$. Under this assumption, it is expected to obtain the global minimizer in the limiting process. Note that the expression (3.39) can be viewed as the counterpart of (3.35)-(3.37).

For comparison purpose, we recall the global smoothed-dual algorithm in [14], where we remain the notation in [14], and $\vec{\mathbf{p}}$ plays the same role as $\vec{\lambda}_1$ in Algorithm 2.

Global Smoothed-Dual (GSD) Algorithm

1. Initialization: set $\vec{\mathbf{p}}^0 = \mathbf{0}$ and choose $\varepsilon, \delta > 0$;
2. For $n = 0, 1, \dots$
 - (i) Compute $\mathbf{v}^n$ by (3.39) with $\mathbf{p}_i^n$ in place of $\mathbf{p}_i$;
 - (ii) Compute $\vec{\mathbf{p}}^{n+1}$ by the projection method:

$$\mathbf{p}_i^{n+1} = \operatorname{Proj}_{\mathcal{S}}(\mathbf{p}_i^n + \delta \nabla v_i^n), \quad 1 \leq i \leq m,$$

where $\operatorname{Proj}_{\mathcal{S}}$ is the projection operator upon the convex set $\mathcal{S}$;

3. Endfor till some stopping rule meets;
4. Set $\vec{\mathbf{p}}^* = \vec{\mathbf{p}}^{n+1}$ and define

$$v_k^* = \begin{cases} 1, & k = \min \{ \arg \min_{1 \leq i \leq m} \{ \operatorname{div} \mathbf{p}_i^* + \lambda f_i \} \}, \\ 0, & \text{otherwise,} \end{cases}$$

for all $1 \leq k \leq m$.

2. A perspective from the level-set method

It is interesting to interpret (3.1) from the perspective of level-set method, where $\{v_i\}$ can be regarded as a piecewise constant interpolation of the phases

(or subregions $\{\Omega_i\}$), that is, v_i takes binary values, i.e., $v_i = 1$, if $\mathbf{x} \in \Omega_i$, and $v_i = 0$ elsewhere. This should be in contrast to the level-set method [97] and its important variant [73] for image segmentation.

The multiphase level set framework in [135] labels the phases by the combinations of the signs of the level-set functions (usually taken as the signed distance functions). Therefore, n level-set functions $\{v_j\}_{j=1}^n$ can label $m = 2^n$ phases, and the corresponding indicator functions can be expressed as the Heaviside functions $H(v_j)$. Notice that the level-set functions should satisfy the Eikonal equation: $|\nabla v_j| = 1$, and this constraint should be taken into account in the implementation. To avoid such a re-initialization, one may penalize this constraint as in [67, 74]. For the purpose of comparison, we briefly review the new level-set method in [67] for (3.1) with $m = 2$. Let v be the signed distance function such that $v > 0$ (resp. $v < 0$) if $\mathbf{x} \in \Omega_1$ (resp. $\mathbf{x} \in \Omega_2$). Then v should satisfy $|\nabla v| = 1$, and the characteristic function of Ω_1 (resp. Ω_2) is $H(v)$ (resp. $1 - H(v)$). The essential idea in [67] is to minimize the penalized energy:

$$\min_v \left\{ E_{\text{Li}}(v) := \lambda \int_{\Omega} g H(v) d\mathbf{x} + \int_{\Omega} \delta(v) |\nabla v| d\mathbf{x} + \frac{\mu}{2} \int_{\Omega} (|\nabla v| - 1)^2 d\mathbf{x} \right\}, \quad \mu > 0, \quad (3.40)$$

where δ is the Dirac delta function, and $g = (f_1 - f_2)/2$. As in [67], we can minimize E_{Li} by evolving the gradient flow (see (3.45) below).

The piecewise constant level-set method (PCLSM) is mentioned in detail in Chapter 1. Then under the setting of (1.28)-(1.31), we replace ψ with v , the model (3.1) is translated to

$$\min_{v \in \{1, \dots, m\}} \left\{ E_P(v) := \lambda \sum_{i=1}^m \int_{\Omega} f_i \chi_i(v) d\mathbf{x} + \sum_{i=1}^m TV(\chi_i(v)) \right\}. \quad (3.41)$$

By imposing the constraint $K_m(v) = 0$, we obtain the primal-dual model (cf. [24]):

$$\min_{K_m=0} \max_{\vec{\mathbf{p}} \in \mathcal{S}} \left\{ F_P(v, \vec{\mathbf{p}}) := \sum_{i=1}^m \langle \chi_i(v), \text{div} \mathbf{p}_i + \lambda f_i \rangle \right\}. \quad (3.42)$$

We see the resemblance between the primal-dual models (3.5) and (3.42). Notice that χ_i is a polynomial of v of degree m . As a result, the reduction using the augmented Lagrangian method similar to the procedure in Subsection 3.1.3 leads to the coupled system, where it is not possible to derive the expression of v in terms of the dual variables. Hence, the algorithm becomes extremely complicated.

For the sake of comparison, we present the counterpart of (3.40) for two-phase case. As described in Subsection 3.1.1, we label Ω_1 (resp. Ω_2) by $v = 1$ (resp. $v = -1$), and then we can formulate (3.1) with $m = 2$ into (3.6). By penalizing the constraint $v^2 = 1$, we have the counterpart of (3.40):

$$\min_v \left\{ E_{\text{PC}}(v) := \lambda \int_{\Omega} g v d\mathbf{x} + \int_{\Omega} |\nabla v| d\mathbf{x} + \frac{\tilde{\mu}}{4} \int_{\Omega} (v^2 - 1)^2 d\mathbf{x} \right\}, \quad \tilde{\mu} > 0, \quad (3.43)$$

where $g = (f_1 - f_2)/2$. One can evolve the gradient flow to minimize (3.43):

$$\frac{\partial v}{\partial t} = \operatorname{div} \left(\frac{\nabla v}{|\nabla v|} \right) + \tilde{\mu}(1 - v^2)v - \lambda g, \quad (3.44)$$

which is closely related to the Allen-Cahn model for phase transitions [3] but with a nonlinear TV-diffusion operator rather than a linear Laplace-diffusion. It was named as the TV-Allen-Cahn model in [136, 110].

Next, we compare the performance of five algorithms: the level-set method in [67] (cf. (3.40)), the PCLSM (cf. (3.43)), GSD Algorithm, and our proposed Algorithms 1-2. Here, we minimize (3.40) based on the code on the “<http://www.engr.uconn.edu/~cmli/research/>” for [67], and also test the algorithms on the image (but with noise) from this site (see Figure 3.1(a) and Figure 3.1(g)). More precisely, the work [67] minimized (3.40) by evolving the gradient flow:

$$v^{n+1} = v^n + \tau_{\text{ls}} \left(\operatorname{div} \left(\frac{\nabla v^n}{|\nabla v^n|} \right) + \frac{\mu}{\delta_{\epsilon}(v^n)} \left(\Delta v^n - \operatorname{div} \left(\frac{\nabla v^n}{|\nabla v^n|} \right) \right) - \lambda g \right), \quad (3.45)$$

where τ_{ls} is the time step size, and the smoothed delta function δ_{ϵ} is defined as in the code for [67]. For ease of comparison, we evolve (3.44) as (3.45):

$$v^{n+1} = v^n + \tau_{\text{pclsm}} \left(\operatorname{div} \left(\frac{\nabla v^n}{|\nabla v^n|} \right) + \tilde{\mu}(1 - (v^n)^2)v^n - \lambda g \right), \quad (3.46)$$

where τ_{pclsm} is the time step size. Here, spatial differential operators are discretized and initial conditions are taken as in [67].

We compare the numerical energy corresponding to the original model (1.13) defined by

$$E(v^n) = \int_{\Omega} |\nabla v^n| d\mathbf{x} + \lambda \int_{\Omega} g v^n d\mathbf{x} \quad \text{with} \quad g = (f_1 - f_2)/2,$$

where v^n is respectively computed by (3.45), (3.46), GSD, Algorithm 1 and Algorithm 2 for the same λ . Correspondingly, we use the stopping rule: $|E(v^{n+1}) - E(v^n)| < \varsigma$. We tabulate in Table 3.1 the iteration number, computational time and numerical energy $E(v^n)$ obtained by five algorithms (terminated with $\varsigma = 10^{-2}$, and we refer to the caption of Figure 3.1 for the computational setup) for two images with different noise levels in Figure 3.1(a) and Figure 3.1(g). We present in Figure 3.1 the final segmented images. We find that at low noise level, all five algorithms produce satisfactory segmentation results with similar numerical energy, but the latter three algorithms converge much faster. However, when we increase the level of noise from 0.001 to 0.1, the former two algorithms fail to converge, while GSD and our proposed algorithms lead to acceptable results, and are quite robust to large noise.

3.2.2 Numerical tests

The previous comparison study demonstrated the advantages of GSD [14] and Algorithms 1-2. Indeed, the recent work [14] conducted a comprehensive comparison of GSD with several popular methods including the Alpha expansion and Alpha-Beta swap [17, 18], the method of Pock et al. [107], and the algorithm in Lellmann et al. [64]. The GSD has exhibited the strength over these algorithms in most comparison tests. In what follows, we put our emphasis on the comparison with GSD, and examine the performance of Algorithms 1-2 over the methods in [17, 18, 64].

Table 3.1: Comparison of five algorithms.

Figure 3.1(a)	iterations	time (second)	energy
Level Set [67]	306	2.0	0.331×10^3
PCLSM (3.43)	415	2.3	0.407×10^3
GSD	17	0.4	0.313×10^3
Alg. 1	11	0.3	0.306×10^3
Alg. 2	14	0.3	0.309×10^3
Figure 3.1(g)	iterations	time (second)	energy
Level Set [67]	1000	5.6	2.32×10^3
PCLSM (3.43)	1000	4.8	1.31×10^3
GSD	72	1.4	0.827×10^3
Alg. 1	40	0.9	0.833×10^3
Alg. 2	53	1.0	0.832×10^3

To set up a relatively sound criterion for comparison, we choose some typical images used in the aforementioned papers for testing the algorithms, unify the choice of the parameters of the model and compare the numerical energy associated with the original model (3.1). Moreover, we use the relative dynamic error:

$$E_{l^1}(\vec{p}^{n+1}, \vec{p}^n) := \frac{\|\vec{p}^{n+1} - \vec{p}^n\|_{l^1}}{\|\vec{p}^{n+1}\|_{l^1}} \leq \eta, \quad (3.47)$$

where $\|\vec{p}\|_{l^1} = \max_i(\sum |\mathbf{p}_i|)$ (with $\vec{\lambda}_1$ in place of $\vec{p}$ for Algorithm 2), and the summation is over all the pixels, for a prescribed tolerance $\eta > 0$ as the stopping rule, for all the comparisons below.

1. Two-class/phase case

As opposed to the GSD algorithm and Algorithm 2 (with $m = 2$), Algorithm 1 only requires to evolve one pair of primal and dual variables. Accordingly, the computational cost can be halved and the algorithm is expected to be more

Figure 3.1: Comparison of five algorithms. Here, we take $\lambda = 5 \times 10^{-5}$, $\tau = 0.1$, $\delta = 0.05$, $\tau_{\text{ls}} = \tau_{\text{pclsm}} = 10^{-5}$, $\epsilon = 1$, $\mu = 1$, $\tilde{\mu} = 10^3$, $\varepsilon = 0.1$ and $f_i = |I - c_i|^2 (i = 1, 2)$. Rows 1 and 2 (a)-(f): noise level $d = 0.001$ and $\beta = 0.01$; Rows 3 and 4 (g)-(l): noise level $d = 0.1$ and $\beta = 0.1$.

stable. Moreover, it can be viewed as a analogy of the Chambolle's dual algorithm [34] (cf. Remark 3.2). Indeed, we shall demonstrate that for the time step size

$\tau \leq 1/8$ (the theoretical prediction by [34]), the method works well, while the global smooth-dual algorithm is relatively restrictive to the time step size, as one will see from all the numerical results below. This is mainly due to the way of computing the primal variable (cf. (3.37) against (3.39)), and updating the dual variables, which is quite different between two approaches. We further remark that Algorithms 1-2 are expected to have a performance similar to the Chambolle's method, but the convergence analysis is much more involved due to the strong nonlinearity (cf. Remark 3.2), which we wish to report in a future work.

Assume that the intensity values c_1 and c_2 are given and $f_i = |c_i - I|^2 (i = 1, 2)$, where I is the intensity of input image as before. The parameters, time step sizes and noise levels for three sets of tests are listed as follows. Recall that λ is the parameter in the continuous Potts model (3.1), and τ, β (resp. δ, ε) are involved in the Algorithms 1-2 (resp. the global smoothed-dual algorithm). We first test the input "UOL" image of size 256×256 with noise level from low to high. Notice that in all tests, the noise is of "Gaussian" type with zero mean and different variance d . We adopt the following setup:

Set 1. Take $\lambda = 10^{-4}, \tau = \delta = 10^{-1}, \beta = 10^{-5}, \varepsilon = 10^{-2}$, and noise level: $d = 0.05$.

Set 2. Take $\lambda = 7 \times 10^{-5}, \tau = 0.1, \delta = 0.05, \beta = 10^{-5}, \varepsilon = 10^{-2}$, and noise level: $d = 0.3$.

Set 3. Take $\lambda = 5 \times 10^{-5}, \tau = 0.1, \delta = 0.01, \beta = 10^{-5}, \varepsilon = 0.05$, and noise level: $d = 0.5$.

We present in Figure 3.2 the input images with noise and segmentation results by three algorithms at the iteration terminated by (3.47) with $\eta = 10^{-2}$. We observe from Figure 3.2 (a)-(d) that when the noise level is low, three algorithms exhibit a similar performance, and converge quite fast. However, when we increase the noise level, Algorithm 1 yields better segmentation (or classification, see the

magnified portion of the letter **U** in Figure 3.2 (m)-(p)), and converges relatively faster. However, the time step size δ in the global smoothed-dual algorithm should be chosen much smaller than τ in Algorithm 1 and Algorithm 2. We also point out the algorithm is not sensitive to the parameter β (refer to Subsection 3.2.2 below for a discussion). We plot the decay of numerical energy corresponding to the original energy functional (3.1), i.e., $E(\mathbf{v}^n)$ in Figure 3.3 for the comparison tests: **Sets 1-3**. Observe that Algorithms 1-2 enjoy a faster decay of energy.

2. Multiphase case

We now turn to the comparison of Algorithm 2 and the global smoothed-dual algorithm for multiphase images. In this case, we take $f_i = |c_i - I|$ with $1 \leq i \leq 4$ for given $\{c_i\}$, and test the four-phase image (cf. Figure 3.4(a)) of size 90×90 with three typical geometric objects and different noise levels. We consider two sets of tests with the following setup:

Set 4. Take $\lambda = 0.2, \beta = 10^{-5}, \tau = 0.1, \delta = 0.05$, and $\varepsilon = 0.1$, and noise level: $d = 0.005$.

Set 5. Take $\lambda = 0.05, \beta = 0.01, \tau = 0.1, \delta = 0.05$ and $\varepsilon = 0.1$, and noise level: $d = 0.05$.

In Figure 3.4, we plot the segmentation results by two algorithms with stopping rule $\eta = 5 \times 10^{-3}$. Once again, we observe the advantages of Algorithm 2. The comparison of numerical energy decay is depicted in Figure 3.5, and we see a faster decay rate for the proposed algorithm. Indeed, we find that Algorithm 2 is robust for noise (in general, we choose the parameter β bigger if the noise level is high), and works for large time step size as the Chambolle's algorithm. In Table 3.2, we tabulate the number of iterations (to meet the stopping rule (3.47) with $\eta = 5 \times 10^{-3}$), the computational time and numerical energy at the last step of iteration, for two algorithms. In fact, the cost for per iteration of two algorithms is almost the same, but Algorithm 2 allows to use a large time step size, so it saves some computational time.

Figure 3.2: Comparison of Algorithms 1-2 and the global smooth-dual (GSD) method for two-phase case. Row 1 (a)-(d) for **Set 1**; Row 2 (e)-(h) for **Set 2**, Row 3 (g)-(l) for **Set 3**, and Row 4 (m)-(p) the magnified portion of the images in Row 2. The iteration is stopped by (3.47) with $\eta = 10^{-2}$. The Algorithms 1-2 converge faster, allow to use larger time step size, and are not sensitive to the choice of the parameter β .

Figure 3.3: Comparison of decay of numerical energy for **Set 1-3**.

Figure 3.4: Comparison of Algorithms 2 and the global smooth-dual algorithm (GSD) for four-phase case. Row 1 (a)-(d) for **Set 4** and Row 2 (e)-(h) for **Set 5**. The iteration is terminated by (3.47) with $\eta = 5 \times 10^{-3}$. Both algorithms work for large noise level. As before, Algorithm 2 converges faster, and produces slightly better results with larger time step.

We next test more multiphase images taken from the papers that we mentioned before, and compare the quality of segmentation with the Alpha expansion and Alpha-Beta swap [17, 18], and the algorithm in Lellmann et al. [64].

Figure 3.5: Comparison of decay of numerical energy for test **Set 4** and **Set 5**.

Table 3.2: Comparison of Algorithm 2 and GSD.

	iterations		time (second)		energy	
	GSD	Alg. 2	GSD	Alg. 2	GSD	Alg. 2
Set 4	72	55	2.3	1.8	1.093×10^4	1.094×10^4
Set 5	103	75	3.6	2.6	7.177×10^3	7.163×10^3

Alpha expansion and Alpha-Beta swap have been regarded as the state of the art for approximately minimizing the discrete version of (1.13) with anisotropic total variation term. Indeed, due to the different nature of algorithms, it is hard to give a very quantified criterion for comparison, so we just depict the segmentation results in Figure 3.6. We find from the first row of Figure 3.6 that the Alpha method and our algorithm give better results than the rest three. We also observe from the second row of Figure 3.6 that the first three algorithms give similar results, while our algorithm produces the best segmentation. From the last row of Figure 3.6, we see that our algorithms and GSD outperform the other methods.

Figure 3.6: Comparison of the Alpha, Alpha-Beta, Lellmann's method, GSD and Algorithm 2.

Figure 3.7: Completion of edges. This shows the proposed algorithm can produce the global optimum.

3. Triple-junction experiments

Next, we test a very typical example relative to triple-junctions, which has been considered by many authors, see, e.g., [144, 104, 64, 37, 14, 75]. The task is to inpaint the edges (or boundaries) of the subregion covered by the disk in Figure 3.7(a) and Figure 3.7(d). It is expected to generate a triple-junction, that is, the completed three edges suppose to form three 120° angles at the junction. In the computation, the data terms f_i inside the disk (white part) are taken to be zero, and $f_i = |c_i - I|^2$ outside the disk, and the parameters in Algorithm 2 are chosen as $\lambda = 10^{-4}$, $\beta = 0.5$ and $\tau = 0.1$. It is important to point out that in this case, the non-smooth problem $\min_{1 \leq i \leq m} \{\operatorname{div} \mathbf{p}_i^* + \lambda f_i\}$ may not have a unique minimum value (see the second case of Theorem 3.2) in the covered subregion, so in theory the algorithm may fail to find the global optimum. However, by slightly increasing the diffusion effect tuned by the parameter β , we are able to obtain a very satisfactory completion of the edges inside the disk. This also indicates that the total variation of the labeling functions (or level-set functions) provides a good characterization of the edge set, and the algorithm produces very accurate approximation, even with large portion of incomplete (or missing) edges (see Figure 3.7(e)), namely, achieve the global optimum. We also quote the result in [75] (obtained by other means), and visualize a good agreement.

4. Selection of the parameters and fidelity terms

Hereafter, we provide some insights into the selection of parameters β and λ , and the fidelity terms f_i . It is seen from the previous numerical tests that the parameter β in Algorithms 1-2 plays an important part for smoothing and regularizing the output images. A natural question is: what is the range of β for which the algorithms perform well? From our experiments, we have found that for two-phase cases, it can be chosen quite flexibly. In principle, the algorithms perform better if we choose larger β for higher level of noise, in particular, for multiphase cases. To further illustrate this, we test Algorithm 2 with different β

on the images with two levels of noise in Figure 3.8(a) and Figure 3.8(g), and fix τ and other parameters. In all cases, the algorithm is terminated by (3.47) with $\eta = 5 \times 10^{-3}$. We record in Figure 3.9 the history of decay of the energy, and plot in Figure 3.8 the segmentation results. We see that for $\beta = 10^{-k}$ with $k = 1, 2, 3$, the algorithm produces similar results with similar amount of iterations, while if β is too small, the convergence rate and quality of segmentation degrade.

Figure 3.8: Results obtained by Algorithms 2 with different β . In the computation, we take $\tau = 0.1$ and $f_i = |c_i - I|(i = 1, \dots, 4)$. Row 1 (a)-(f): noise level $d = 0.01$ and $\lambda = 0.2$; Row 2 (g)-(l): noise level $d = 0.02$ and $\lambda = 0.1$. The iteration is terminated by (3.47) with $\eta = 5 \times 10^{-3}$, and we indicate the number of iterations and the values of β (in the parentheses).

Figure 3.9: Decay of numerical energy for different β .

We next discuss the sensitivity to the parameter λ , which is born with the model (3.1) with $f_i = |c_i - I|$. In Figure 3.10 (a)-(f), we fix the noise level d , and the parameters τ, β , but choose different λ ranging from 0.01 to 0.2, while in Figure 3.10 (g)-(l), we vary the noise level, but fix $\lambda = 0.1$ in Algorithm 2. We see that Algorithm 2 enjoys a very similar performance in all tests. Indeed, the choice of λ is relatively flexible from our various experimental experiences. Some delicate analysis and adaptive determination of λ for the ROF-type model can be found in e.g., [97, 50], where λ is known as a scale variable. However, this requires significant computational cost to adapt λ .

Finally, we give some observations on the choice of the fidelity terms $f_i = |c_i - I|^p$ ($p = 1, 2$). We remark that some in-depth analysis for TV-regularized L^p model can be found in, e.g., [39, 22, 120]. In general, the quadratic case ($p = 2$) is computationally easier to handle and more desirable for Gaussian noise removal, while the l_1 -norm ($p = 1$) is more capable of preserving contrast. In this context, $\{c_i\}$ are preassigned, and we realize that the use of $p = 1$ or 2 largely affects the choice of the parameter λ . Observe that

$$\lambda_2 \int_{\Omega} |c_i - I|^2 v_i d\mathbf{x} \sim \lambda_2 \|c_i - I\|_{\xi \in \Omega} \int_{\Omega} |c_i - I| v_i d\mathbf{x},$$

Figure 3.10: Results obtained by Algorithms 2 with fixed $\tau = 0.1$, $\beta = 0.01$. Row 1: noise level $d = 0.1$ for all tests, but different λ . Row 2: $\lambda = 0.1$ for all tests, but noise level $d = 0.005, 0.1, 0.2$ from left to right. The iteration is terminated by (3.47) with $\eta = 5 \times 10^{-3}$.

where for clarity, we set $\lambda = \lambda_p$ for $p = 1, 2$. For piecewise constant image I , we can empirically choose $\lambda_1 = \max_{i \neq j} |c_i - c_j| \lambda_2$. In Figure 3.11, we set $\lambda_1 = 200 \lambda_2$, and see that Algorithm 2 for two models work almost equally well.

Figure 3.11: Results obtained by Algorithms 2 with different $f_i = |c_i - I|, |c_i - I|^2$ and $\lambda_1 = 200\lambda_2$. Here, $\tau = 0.1, \beta = 0.01$. (a)-(c): $\lambda_1 = 0.2$; (d)-(f): $\lambda_1 = 0.1$. The iteration is terminated by (3.47) with $\eta = 5 \times 10^{-3}$.

5. Applications to image segmentation

Finally, we apply the methods to the piecewise constant Mumford-Shah model (1.12), for which $\{c_i\}$ are unknowns computed from the mean values:

$$c_i = \frac{\int_{\Omega} I v_i d\mathbf{x}}{\int_{\Omega} v_i d\mathbf{x}}, \quad 1 \leq i \leq m. \quad (3.48)$$

For fixed c_i , the model (1.12) is a special case of (1.13) with $f_i = |c_i - I|^2$. Therefore, with a slight modification of Algorithm 2, we can obtain the following algorithm for the multiphase piecewise constant image segmentation.

Algorithm 3

1. Initialization: set $\vec{\lambda}_1^0 = \mathbf{0}$ and choose $\beta, \tau > 0$, and $\mathbf{c}^0$;
2. For $n = 0, 1, \dots$
 - (i) Compute λ_3^n by (3.36) with $f_i^n = |c_i^n - I|^2$ in place of f_i ;
 - (ii) Compute $\mathbf{v}^n$ by (3.37) with $h_i = \operatorname{div} \lambda_{1i}^n + \lambda f_i^n$ and $\lambda_3 = \lambda_3^n$;

(iii) Compute $\vec{\lambda}_1^{n+1}$ by

$$\lambda_{1i}^{n+1} = \frac{\lambda_{1i}^n - \tau \nabla v_i^n}{1 + \tau |\nabla v_i^n|}, \quad 1 \leq i \leq m;$$

(iv) Compute $\mathbf{c}^{n+1}$ by

$$c_i^{n+1} = \frac{\int_{\Omega} I v_i^n d\mathbf{x}}{\int_{\Omega} v_i^n d\mathbf{x}}, \quad 1 \leq i \leq m.$$

3. Endfor till some stopping rule meets;

4. Set $\vec{\lambda}_1^* = \vec{\lambda}_1^{n+1}$, $\mathbf{c}^* = \mathbf{c}^{n+1}$ and define

$$v_k^* = \begin{cases} 1, & k = \min \{ \arg \min_{1 \leq i \leq m} \{ \operatorname{div} \lambda_{1i}^* + \lambda f_i^* \} \}, \\ 0, & \text{otherwise,} \end{cases}$$

for all $1 \leq k \leq m$, where $f_i^* = |c_i^* - I|^2$.

Similarly, the global smoothed-dual algorithm can be modified for the segmentation problem (see Algorithm 2 in [14]), and for clarity, we term it as GSD-Seg in short.

We first compare Algorithm 3 with GSD-Seg for segmenting the “desk” image in Figure 3.12(a) of size 205×255 containing three typical geometric objects with three phases. For this example, it is not an easy task to segment the faces of the objects, for instance, the two faces of the tetrahedron. In the test, both algorithms are terminated by (3.47) with $\eta = 5 \times 10^{-3}$ (refer to the caption of Figure 3.12 for the selection of parameters and time step size). As before, both algorithms produce quite satisfactory and accurate segmentation, while Algorithm 3 converges relatively faster. We also tabulate in Table 3.3, the number of iterations (to meet the stopping rule (3.47) with $\eta = 5 \times 10^{-3}$), the computational time and numerical energy at the last step of iteration, for two algorithms. Once again, the cost for per iteration of two algorithms is almost the same, but Algorithm 3 allows to use a large time step size.

Figure 3.12: Comparison of Algorithm 3 and GSD-Seg, where $\tau = 0.1$, $\delta = 0.02$, $f_i = |c_i - I|^2$ and other parameters: (a)-(c) (three phases of size 205×255): $\lambda = 5 \times 10^{-4}$, $\beta = 10^{-2}$, $\varepsilon = 0.1$. Given this noise input image, it is challenging to segment the faces of the objects. The iteration of both algorithms is terminated by (3.47) with $\eta = 5 \times 10^{-3}$. Algorithm 3 slightly outperforms GSD-Seg.

Table 3.3: Comparison of Algorithm 3 and GSD-Seg.

iterations		time (second)		energy	
GSD-Seg	Alg. 3	GSD-Seg	Alg. 3	GSD-Seg	Alg. 3
115	87	15	11	1.320×10^4	1.314×10^4

We next test some more images which are used as examples in most of the papers that we mentioned before. We refer to the caption of Figure 3.13 for the choices of the parameters in Algorithm 3. In all cases, the algorithm is terminated by (3.47) with $\eta = 5 \times 10^{-3}$. We see that in all tests, Algorithm 3 provides quite satisfactory segmentation and converges fast.

Figure 3.13: More tests on Algorithm 3, where $\tau = 0.1$, $f_i = |c_i - I|^2$ and other parameters: (a)-(b) (five phases): $\lambda = 10^{-4}$, $\beta = 10^{-3}$, size: 150×150 ; (c)-(d) (five phases): $\lambda = 10^{-3}$, $\beta = 10^{-2}$, size: 90×90 ; (e)-(f) (four phases): $\lambda = 10^{-3}$, $\beta = 10^{-5}$, size: 90×90 ; (g)-(h) (three phases): $\lambda = 10^{-3}$, $\beta = 10^{-2}$, size 321×481 ; (i)-(j) (three phases): $\lambda = 10^{-3}$, $\beta = 10^{-2}$, size: 167×250 and (k)-(l) (five phases): $\lambda = 10^{-3}$, $\beta = 10^{-2}$, size: 512×512 . The iteration is terminated by (3.47) with $\eta = 5 \times 10^{-3}$.

Robust Edge Detection Using A Modified Mumford-Shah Model

Edge detection plays an immensely important role not only in image processing and computer vision, but also in many other fields, e.g., material science and physics [134, 16]. In the context of image processing, edge detection is to find the boundaries of objects in a digital image. Many methods have been developed for this purpose. Classical approaches manipulating image gradient [29, 48, 122] are limited, which is largely due to the presence of noise that usually deteriorates the accuracy of edge identification. Another important class of methods is based on anisotropic diffusive equations [32, 4, 102]. In general, such models are designed to prevent smoothing near the edges but to encourage diffusion in the homogeneous regions. However, they may face difficulties in dealing with low signal to noise ratios (SNRs). Note that some reaction and diffusion equations [126, 139], edge flow [80], snake/active contour models [62, 31, 13] and universal gravity [125] have been proposed for edge detection. Recent methods use a variety of filter banks to improve the accuracy of detection and the interested readers are referred to [23, 85, 76, 105, 131].

In distinctive contrast with (e.g., region-based) image segmentation, edge detection is also concerned with finding open curves. In principle, an open-ended

curve can be expressed by a parametric form, but it is not a natural representation for evolution. Among a few techniques for manipulating open curves, Leung and Zhao [66] proposed a grid based particle method to represent an open curve. We also refer to [121, 15] for different ideas. It is known that the level-set method [98] has become a versatile tool for various applications involving curve/surface evolution. However, it is suitable to deal with a closed curve that separates a region into interior and exterior parts.

The idea in the recent conference report [137] is to embed an open (or a closed) curve of interest in a narrow region (or band) with the curve being part of the (one-sided) boundary. From geometric point of view, such a region is formed by the parallel curve (also known as the offset curve) [133]. This allows to integrate a binary level-set method with the Mumford-Shah (MS) model (see (1.1) in Chapter 1). It should be pointed out that only very preliminary algorithms were proposed in the note [137] for this modified model. On the other hand, by viewing an edge as a narrow surrounding band, the global algorithm proposed in Chapter 3 is not applicable to this limiting case. With this in mind, the purposes of this chapter are to introduce two fast algorithms using the proximity algorithm [90, 91] and split Bregman method [59] for solving the subproblem of the level-set function. Indeed, the derivation of the algorithms from sub-differentiation can mathematically handle the limiting case. We also intend to compare the performance of the algorithms with the methods based on the modified Mumford-Shah model by Ambrosio and Tortorelli (AT) [5, 6] for edge detection. We demonstrate that the proposed methods outperform the AT's approach in terms of ease of implementation, quality of detection, and robustness to noise.

This chapter is organized as follows. A brief review of the model [137] is given in Section 4.1. In Section 4.2, we describe the algorithms for solving the minimization problem which mainly include proximity algorithm, split Bregman algorithm and the Fourier method. In Section 4.3, a comparison study is conducted as well as the numerical results.

4.1 The Modified Mumford-Shah Model

We start with recalling the full Mumford-Shah model (cf. (1.1) in Chapter 1):

$$\min_{u, \Gamma} \left\{ E(u, \Gamma) = \mu \int_{\Omega \setminus \Gamma} |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} (u - I)^2 d\mathbf{x} + |\Gamma| \right\}, \quad (4.1)$$

where I is a given image on an open bounded domain $\Omega \in \mathbb{R}^2$, u is the piecewise smooth approximation of I , $|\Gamma|$ is the length of the edges spread all over the image domain, and μ, ν are positive parameters to balance the three terms in (4.1).

An important issue here is how to characterize the edges. Let $\mathbf{r} \in \Gamma$ be a curve with the parameterization: $\mathbf{r}(t) = (x(t), y(t)), t \in [0, 1]$. Without loss of generality, assume that $\mathbf{r}(t)$ is regular (i.e., $|\mathbf{r}'(t)| \neq 0$ for all $t \in (0, 1)$), and it has finite length and curvature. Then the parallel or offset curve generated by $\mathbf{r}(t)$ is defined by (cf. [54, 133]):

$$\mathbf{r}_d(t) = \mathbf{r}(t) + d\mathbf{n}(t), \quad \forall t \in [0, 1], \quad (4.2)$$

where $\mathbf{n}(t)$ is the unit normal to $\mathbf{r}(t)$ at each point, and d is a preassigned signed distance. This defines a positive (exterior, $d > 0$) or negative (interior, $d < 0$) offset (see some examples in Figure 4.1). Let us denote the total length of $\mathbf{r}$ and $\mathbf{r}_d$ by L and L_d , respectively. Then we have (cf. Lemma 3.1 in [54]):

$$L_d = \int_0^1 |1 + \kappa d| |\mathbf{r}'| dt = \int_0^L |1 + \kappa d| ds = |L + d \Delta\theta|,$$

where κ is the curvature and $\Delta\theta$ is the total angle of rotation of the normal $\mathbf{n}$ to $\mathbf{r}$ between $t = 0$ and $t = 1$, measured in the right-handed sense. Moreover, the area, denoted by A_d , between the generator $\mathbf{r}$ and the offset $\mathbf{r}_d$ is given by (see [54] again):

$$A_d = \frac{1}{2}(L + L_d)|d|.$$

We see that the curve $\mathbf{r}$ is immersed in the (closed) band (with area A_d) denoted by R_d .

Define the binary level-set function: $\psi = 1$ if $\mathbf{x} \in R_d$, while $\psi = -1$ otherwise. Recall that the total variation of the indicator function $\chi_{R_d} = (1 + \psi)/2$

Figure 4.1: The closed narrow region R_d formed by the curve $\Gamma : \mathbf{r}(t)$ and its exterior parallel curve (dashed line) for closed curves and open curves (the dotted lines connected the corresponding starting points and end-points of the curve Γ and its exterior parallel curve, respectively), where the dot-dashed line is the interior parallel curve. The last one is an example of an intersected curve that can be split into simple open or closed curves.

characterizes the perimeter of R_d (see (1.7)), so we deduce from (4.2) that for $0 < d \ll 1$,

$$TV(\chi_{R_d}) = 2L + O(d) \quad \Rightarrow \quad TV(\psi) \approx 4L.$$

Collecting all these leads to the approximation of the MS model (4.1), dubbed as the modified MS model:

$$\min_{u, \psi^2=1} \left\{ E_{MMS}(u, \psi) = \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + TV(\psi) \right\}, \quad (4.3)$$

where the constraint $\psi^2 = 1$ is imposed for the binary level set as in Chapter 1 (see (1.24)).

We see from the above that the essence lies in the characterization of the edge set Γ and its total length $|\Gamma|$ in the MS model (4.1) by using the narrow surrounding band and the binary level-set method. This situation is reminiscent to the way of treating Γ and approximating $|\Gamma|$ in Ambrosio and Tortorelli [5, 6], where the basic idea is to associate Γ with the jump Γ_u of the unknown u . The

approximation of (4.1) in [5, 6] is given by

$$\begin{aligned} E_{AT}(u, v) = & \mu \int_{\Omega} (v^2 + o_{\varepsilon}) |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} (u - I)^2 d\mathbf{x} \\ & + \int_{\Omega} \left(\varepsilon |\nabla v|^2 + \frac{1}{4\varepsilon} (v - 1)^2 \right) d\mathbf{x}, \end{aligned} \quad (4.4)$$

where ε is a sufficient small parameter, and o_{ε} is any non-negative infinitesimal quantity approaching 0 faster than ε . Here, v plays a role similar to ψ in (4.3), that is, $v \approx 0$ if $x \in \Gamma_u$, and $v \approx 1$ otherwise. A rigorous analysis (see, e.g., [5, 19]) shows that the last term converges to the length of $|\Gamma_u|$ in Γ -convergence sense [46]. The width of transition from $v = 0$ to 1 is about $O(\varepsilon)$. This Ambrosio and Tortorelli approximation (4.4) has been extensively studied and extended for segmentation and other applications (see, e.g., [116, 82, 132, 56, 117]).

4.2 Minimization Algorithm

This section is to introduce the minimization algorithms for solving (4.3), which mainly consist of the proximity algorithm and split Bregman method for resolving the binary level-set function ψ , and the Fourier methods for finding u in (4.3).

Although the modified MS model (4.3) looks analogous to the two-phase segmentation model studied in Chapters 2-3, we face the situation that the edge set $\Gamma = \{\psi = 1\}$ could be of measure zero in the limiting process. It is questionable to define the dual formulation (see (3.7)). For this reason, we resort to the proximity algorithm and split Bregman method, where the use of subdifferential can bypass this.

It is clear that the set $\{\psi^2 = 1\}$ is nonconvex. We first relax it to a convex set:

$$K = \{\psi \in BV(\Omega) : |\psi| \leq 1\},$$

where $BV(\Omega)$ is the space of functions of bounded variation (cf. (1.6)). We

further denote the characteristic function of the convex set K by

$$l_K(\psi) := \begin{cases} 0, & \psi \in K, \\ +\infty, & \psi \notin K. \end{cases}$$

Then, we consider the unconstrained minimization problem derived from (4.3):

$$\min_{u, \psi} \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + TV(\psi) + \int_{\Omega} l_K(\psi) d\mathbf{x} \right\}. \quad (4.5)$$

Due to the involvement of the last term, it appears difficult to work on the functional directly. Here, we follow the idea in Chapter 2 (see (2.7)-(2.8)) by introducing an auxiliary $q = \psi$ to handle the last term and treat the constraint $q = \psi$ by a penalty method. Thus, the equation (4.5) can be approximated by

$$\begin{aligned} \min_{u, \psi, q} \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} + TV(\psi) \right. \\ \left. + \int_{\Omega} l_K(q) d\mathbf{x} + \frac{r}{2} \int_{\Omega} (q - \psi)^2 d\mathbf{x} \right\}. \end{aligned} \quad (4.6)$$

Using the alternating optimization technique, we split (4.6) into three sub-problems:

- u -subproblem: for fixed ψ , we solve

$$\min_u \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u - I|^2 d\mathbf{x} \right\}. \quad (4.7)$$

- ψ -subproblem: for fixed u and q , we solve

$$\min_{\psi} \left\{ \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + TV(\psi) + \frac{r}{2} \int_{\Omega} (q - \psi)^2 d\mathbf{x} \right\}. \quad (4.8)$$

- q -subproblem: for fixed ψ , we solve

$$\min_q \left\{ \int_{\Omega} l_K(q) d\mathbf{x} + \frac{r}{2} \int_{\Omega} (q - \psi)^2 d\mathbf{x} \right\}. \quad (4.9)$$

One can verify readily that the solution of (4.9) can be explicitly expressed as

$$q = \psi \min \left\{ 1, \frac{1}{|\psi|} \right\}.$$

Next, we present the algorithms for (4.7) and (4.8).

4.2.1 FFT for solving u

We first consider (4.7). Notice that for fixed $\psi \in K$, the functional in (4.7) is convex, so it admits a unique minimizer. The corresponding Euler-Lagrange equation takes the form:

$$\begin{cases} -2\mu \operatorname{div}((1-\psi)^2 \nabla u) + \nu(u - I) = 0, & \text{in } \Omega, \\ \frac{\partial u}{\partial \mathbf{n}} \Big|_{\partial\Omega} = 0, \end{cases} \quad (4.10)$$

where $\mathbf{n}$ is the unit outer normal vector to $\partial\Omega$ as before. Notice that the expected value of ψ is -1 in the homogeneous region. To avoid solving the elliptic problem with variable coefficient, we rewrite the first equation in (4.10) as

$$-8\mu\Delta u + \nu u = 2\mu \operatorname{div}((1-\psi)^2 \nabla u) - 8\mu\Delta u + \nu I := L(u).$$

Then we construct the following iterative scheme

$$\begin{cases} -8\mu\Delta u^{k+1} + \nu u^{k+1} = L(u^k), & \text{in } \Omega, \\ \frac{\partial u^{k+1}}{\partial \mathbf{n}} \Big|_{\partial\Omega} = 0, & k = 0, 1, \dots \end{cases} \quad (4.11)$$

With this treatment, we only need to solve a (modified) Helmholtz equation at each iteration. Then we can use Fourier methods with fast Fourier transform (FFT) to solve it very efficiently (see, e.g., [118]). Such a preconditioning technique is mimic to that e.g., in [33], which enjoys a fast rate of convergence, as $(1-\psi)^2 \approx 4$, a.e. in Ω . The discretization of the differential operators can be found in [34].

4.2.2 Proximity algorithm for solving ψ

Now, we turn to (4.8). We start with introducing some basics of the proximity operator and proximity algorithm. Here, we adopt the setting as in [89, 88, 90, 91], and refer to the references therein for more details.

Definition 4.1. Let H be a real Hilbert space, and h be a convex functional on H , not identically equal to ∞ . The sub-differential of h at $x \in H$ is the set defined by

$$\partial h(x) := \{y \in H : h(z) \geq h(x) + \langle y, z - x \rangle, \forall z \in H\},$$

where $\langle \cdot, \cdot \rangle$ is the inner product of H .

We remark that any $y \in \partial h(x)$ is called a subgradient of h at x . The sub-differential reduces to the classical gradient if $h(x)$ is differentiable.

The definition of the proximity operator introduced by Moreau [90, 91] is recalled below.

Definition 4.2. Let h and H be the same as in Definition 4.1. Then the proximity operator of h at $x \in H$ is defined by

$$\text{prox}_h(x) := \arg \min_{u \in H} \left\{ \frac{1}{2} \|u - x\|_2^2 + h(u) \right\},$$

where $\|v\|_2 = \sqrt{\langle v, v \rangle}$.

A very important property is that the subdifferential of a convex function can be characterized by its proximity operator. We refer to [90, 91] for the detailed proof.

Lemma 4.1. Let h and H be the same as in Definition 4.1. Then for any $x \in H$, we have

$$y \in \partial h(x) \text{ if and only if } x = \text{prox}_h(x + y).$$

Next, we apply Lemma 4.1 to solve the subproblem (4.8). For this purpose, we define

$$\rho(\psi) = \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + \frac{r}{2} \int_{\Omega} (q - \psi)^2 d\mathbf{x}. \quad (4.12)$$

Then (4.8) is equivalent to

$$\min_{\psi} \{J(\psi) = (g \circ B)(\psi) + \rho(\psi)\}, \quad (4.13)$$

where $(g \circ B)(\psi) = TV(\psi)$ with the understanding that

$$g(\cdot) = \|\cdot\|_{L^1(\Omega)}, \quad B = \nabla. \quad (4.14)$$

Then by using the following result, which can be proved in a fashion similar to Proposition 1 in [89], we can obtain the solution of (4.13).

Lemma 4.2. *Let g and B be the same as above, and ρ be a convex functional, e.g., in (4.12). If $\psi \in H(= L^2(\Omega))$ is a solution of (4.13), then for any $\alpha, \beta > 0$, there exists $\mathbf{s} \in H \times H$ such that*

$$\psi = \text{prox}_{\frac{1}{\alpha}\rho}\left(\psi - \frac{\beta}{\alpha}B^*\mathbf{s}\right), \quad (4.15)$$

and

$$\mathbf{s} = (\mathbb{I} - \text{prox}_{\frac{1}{\beta}g})(B\psi + \mathbf{s}), \quad (4.16)$$

where B^* is the adjoint operator of B , and $\mathbb{I}$ is the identity operator.

Conversely, if there exist $\alpha, \beta > 0$, $\mathbf{s} \in H \times H$, and $\psi \in H$ satisfying (4.15) and (4.16), then ψ is a solution of model (4.13).

Proof. We assume that $\psi \in H$ is a solution of the model (4.13). By the Fermat rule in convex analysis, we have

$$\partial J(\psi) = 0.$$

By the chain rule, we obtain

$$\partial(g \cdot B)(\psi) = B^*\partial g(B\psi).$$

It is equivalent to the condition that

$$0 = \partial\rho(\psi) + B^*\partial g(B\psi). \quad (4.17)$$

For $\alpha, \beta > 0$, we can choose a vector $a = \frac{1}{\alpha}\partial\rho(\psi)$ and a vector $s = \frac{1}{\beta}\partial g(B\psi)$ such that

$$0 = \alpha a + \beta B^*s. \quad (4.18)$$

By Lemma 4.1, from the conclusion $s = \frac{1}{\beta} \partial g(B\psi)$ we get (4.16) and from the conclusion $a = \frac{1}{\alpha} \partial \rho(\psi)$ we have that

$$\psi = \text{prox}_{\frac{\alpha}{\rho}}(\psi + a). \quad (4.19)$$

Using (4.17) we conclude that $a = -\frac{\beta}{\alpha} B^* s$ and substitute it into (4.19) we obtain (4.15).

Conversely, suppose that there exist $\alpha, \beta > 0$, $s \in \mathbb{R}^m$, and $\psi \in \mathbb{R}^d$ satisfying (4.15) and (4.16). Again, by Lemma 4.1, (4.15) and (4.16) ensure that $-\frac{\beta}{\alpha} B^* s = \frac{1}{\alpha} \partial \rho(\psi)$ and $s = \frac{1}{\beta} \partial g(B\psi)$, respectively. Combining these two conclusion relations yields that

$$0 = \alpha \left(-\frac{\beta}{\alpha} B^* s \right) + \beta B^* s = \partial \rho(\psi) + B^* \partial g(B\psi).$$

This in turn implies that ψ is a solution of the model (4.13). $\square$

Equipped with Lemma 4.2, we are able to derive the proximity algorithm for (4.13). By Definition 4.2 and (4.15), we obtain

$$\psi = \text{prox}_{\frac{1}{\alpha} \rho} \left(\psi - \frac{\beta}{\alpha} B^* s \right) = \arg \min_{\omega} \left\{ \frac{1}{\alpha} \rho(\omega) + \frac{1}{2} \left\| \omega - \left(\psi - \frac{\beta}{\alpha} B^* s \right) \right\|_{L^2(\Omega)}^2 \right\}. \quad (4.20)$$

In view of (4.12), we find the Euler-Lagrangian equation:

$$\omega - M_1 + \frac{1}{\alpha} \left\{ -2\mu(1 - \omega) |\nabla u|^2 + r(\omega - q) \right\} = 0,$$

where

$$M_1 = \psi - \frac{\beta}{\alpha} B^* s.$$

This implies

$$\omega = \frac{rq + 2\mu |\nabla u|^2 + \alpha M_1}{\alpha + r + 2\mu |\nabla u|^2} \stackrel{(4.20)}{=} \psi. \quad (4.21)$$

It remains to find s in (4.16). The special form of the functional g enables us to formulate the subdifferential and proximity operators in a closed form (see, e.g., [88]):

$$s = (B\psi + s) - \max \left\{ |B\psi + s| - 1/\beta, 0 \right\} \cdot \text{sign}(B\psi + s). \quad (4.22)$$

We summarize the full algorithm for (4.6), called FFT-Proximity (FFT-P) Algorithm, as follows.

FFT-P Algorithm

1. Initialization: set $\mathbf{s}^0 = \mathbf{0}$, and input $\psi^0, u^0, \mu, \nu, \alpha, \beta, r$.
 2. For $n = 0, 1, \dots$,
 - (i) Update u^{n+1} using FFT and the iteration scheme (4.11) with initial value u^n and ψ^n in place of ψ ;
 - (ii) Update q^n by

$$q^n = \begin{cases} \psi^n \min \left\{ 1, \frac{1}{|\psi^n|} \right\}, & \text{if } \psi^n \neq 0, \\ 0, & \text{if } \psi^n = 0; \end{cases}$$
 - (iii) Update ψ^{n+1} by

$$\psi^{n+1} = \frac{rq^n + 2\mu|\nabla u^{n+1}|^2 + \alpha(\psi^n - \frac{\beta}{\alpha}B^*\mathbf{s}^n)}{\alpha + r + 2\mu|\nabla u^{n+1}|^2};$$
 - (iv) Update $\mathbf{s}^{n+1}$ by

$$\mathbf{s}^{n+1} = (B\psi^{n+1} + \mathbf{s}^n) - \max \{ |B\psi^{n+1} + \mathbf{s}^n| - 1/\beta, 0 \} \cdot \text{sign}(B\psi^{n+1} + \mathbf{s}^n).$$
 3. Endfor till some stopping rule meets.
-

Some remarks are in order.

- In this context, the operators $B = \nabla$ and $B^* = -\text{div}$. We refer to [34] for their discretizations.
- It is seen that under the proximal framework and with FFT implementation, the algorithm appears simple to implement. While the proximity algorithm introduces two additional parameters α and β , the algorithm is robust as to be demonstrated in the numerical examples, and some useful guidelines in [89].

4.2.3 Split Bregman method for solving ψ

As already mentioned, another way to get around the imposition of total variation over the edge set Γ (with “zero” measure in the limiting case) is to use the split Bregman method [59], which basically integrates the notion of Bregman distance [20] with augmented Lagrangian techniques (cf. [129]). Indeed, this method has been widely used in image processing (see, e.g., [28, 58, 115, 145] and the references therein).

We start with introducing an auxiliary variable $\mathbf{d}$ and rewrite (4.13) as

$$\min_{\psi} \left\{ \int_{\Omega} |\mathbf{d}| d\mathbf{x} + \rho(\psi) \right\} \quad \text{subject to} \quad \mathbf{d} = (d_1, d_2) = \nabla \psi,$$

where $|\mathbf{d}| = \sqrt{d_1^2 + d_2^2}$ and $\rho(\psi)$ is defined in (4.12). Following [59], we formulate the split Bregman iteration as

$$(\psi^{n+1}, \mathbf{d}^{n+1}) = \arg \min_{\psi, \mathbf{d}} \left\{ \int_{\Omega} |\mathbf{d}| d\mathbf{x} + \rho(\psi) + \frac{\lambda}{2} \int_{\Omega} (\mathbf{d} - \nabla \psi - \mathbf{b}^n)^2 d\mathbf{x} \right\}, \quad (4.23)$$

for given $\mathbf{b}^n$, and

$$\mathbf{b}^{n+1} = \mathbf{b}^n + (\nabla \psi^{n+1} - \mathbf{d}^{n+1}). \quad (4.24)$$

It is equivalent to writing (4.23)-(4.24) in the component form of $\mathbf{d} = (d_1, d_2)$ and $\mathbf{b} = (b_1, b_2)$. This yields

$$\begin{aligned} (\psi^{n+1}, d_1^{n+1}, d_2^{n+1}) = \arg \min_{\psi, d_1, d_2} & \left\{ \int_{\Omega} \sqrt{d_1^2 + d_2^2} d\mathbf{x} \right. \\ & + \mu \int_{\Omega} (1 - \psi)^2 |\nabla u|^2 d\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} (d_1 - \partial_x \psi - b_1^n)^2 d\mathbf{x} \\ & \left. + \frac{\lambda}{2} \int_{\Omega} (d_2 - \partial_y \psi - b_2^n)^2 d\mathbf{x} + \frac{r}{2} \int_{\Omega} (q - \psi)^2 d\mathbf{x} \right\}, \end{aligned} \quad (4.25)$$

and

$$b_1^{n+1} = b_1^n + (\partial_x \psi^{n+1} - d_1^{n+1}), \quad b_2^{n+1} = b_2^n + (\partial_y \psi^{n+1} - d_2^{n+1}).$$

The Euler-Lagrangian equation of (4.25) for ψ with fixed d_1 and d_2 is

$$-\lambda \Delta \psi + 2\mu |\nabla u|^2 (\psi - 1) + r(\psi - q) - \lambda \partial_x (d_1 - b_1^n) - \lambda \partial_y (d_2 - b_2^n) = 0,$$

with the Neumann boundary condition $\frac{\partial \psi}{\partial \mathbf{n}} = 0$.

Once again, this problem can be solved by FFT (cf. Subsection 4.2.1), but in this case, we face the variable coefficient in front of ψ :

$$\begin{cases} -\lambda \Delta \psi^{k+1} + r \psi^{k+1} = F(\psi^k), & \text{in } \Omega, \\ \frac{\partial \psi^{k+1}}{\partial \mathbf{n}} \Big|_{\partial \Omega} = 0, & k = 0, 1, \dots, \end{cases} \quad (4.26)$$

where

$$F(\psi) = -2\mu |\nabla u|^2 \psi + 2\mu |\nabla u|^2 + r q + \lambda \partial_x (d_1 - b_1^n) + \lambda \partial_y (d_2 - b_2^n).$$

For fixed ψ , the optimality condition of (4.25) with respect to d_1 and d_2 gives

$$d_1 = \max \left\{ h - \frac{1}{\lambda}, 0 \right\} \frac{h_1}{h}, \quad d_2 = \max \left\{ h - \frac{1}{\lambda}, 0 \right\} \frac{h_2}{h}, \quad (4.27)$$

where

$$h_1 = \partial_x \psi + b_1^n, \quad h_2 = \partial_y \psi + b_2^n, \quad h = \sqrt{h_1^2 + h_2^2}.$$

Now, we are in a position to present the algorithm. We replace Steps (iii) and (iv) in FFT-P algorithm by the split Bregman method and summarize the full algorithm, termed as FFT-Split Bregman (FFT-SP) Algorithm.

FFT-SP Algorithm

1. Initialization: set $d_1^0 = d_2^0 = b_1^0 = b_2^0 = 0$, and input $\psi^0, u^0, \mu, \nu, \lambda, r$.
2. For $n = 0, 1, \dots$,
 - (i) Update u^{n+1} using FFT and the iteration scheme (4.11) with initial value u^n and ψ^n in place of ψ ;
 - (ii) Update q^n by

$$q^n = \begin{cases} \psi^n \min \left\{ 1, \frac{1}{|\psi^n|} \right\}, & \text{if } \psi^n \neq 0, \\ 0, & \text{if } \psi^n = 0; \end{cases}$$

- (iii) Update ψ^{n+1} using FFT and the iteration scheme by (4.26) with initial value ψ^n , u^{n+1} in place of u and q^n in place of q ;
- (iv) Update d_1^{n+1} and d_2^{n+1} by (4.27) with ψ^{n+1} in place of ψ ;
- (v) Update b_1^{n+1} and b_2^{n+1} by

$$b_1^{n+1} = b_1^n + \left(\partial_x \psi^{n+1} - d_1^{n+1} \right), \quad b_2^{n+1} = b_2^n + \left(\partial_y \psi^{n+1} - d_2^{n+1} \right).$$

3. Endfor till some stopping rule meets.

We point out that due to the use of the Bregman method, this algorithm provides an efficient way for handling the TV-term.

4.3 Numerical Experiments

In this section, we show the performance of the proposed model and algorithms by testing them on some typical images and comparing with some relevant methods.

4.3.1 Algorithm for the Ambrosio and Tortorelli model

As aforementioned, the Ambrosio and Tortorelli (AT) model [5, 6] is somehow mimic to our model in terms of approximating the length of the edge set. Hence, we feel compelled to make a comparison.

The Euler-Lagrange equations of (4.4) read

$$\begin{cases} u = \frac{2\mu}{\nu} \operatorname{div}[(v^2 + o_\epsilon) \nabla u] + I, & \text{in } \Omega, \\ v \left(\frac{1}{4\epsilon} + \mu |\nabla u|^2 \right) = \epsilon \Delta v + \frac{1}{4\epsilon}, & \text{in } \Omega, \\ \frac{\partial u}{\partial \mathbf{n}} = 0, & \text{on } \partial\Omega, \\ \frac{\partial v}{\partial \mathbf{n}} = 0, & \text{on } \partial\Omega. \end{cases} \quad (4.28)$$

We next solve the system (4.28) by the gradient descent method and FFT:

$$\begin{cases} u_t = \frac{2\mu}{\nu} \operatorname{div}[(v^2 + o_\epsilon) \nabla u] - u + I, & \text{in } \Omega, \\ v_t = \epsilon \Delta v - v \left(\frac{1}{4\epsilon} + \mu |\nabla u|^2 \right) + \frac{1}{4\epsilon}, & \text{in } \Omega. \end{cases}$$

We adopt the semi-implicit scheme:

$$\frac{u^{n+1} - u^n}{\Delta t} - \frac{2\mu}{\nu} \Delta u^{n+1} + u^{n+1} = \frac{2\mu}{\nu} \operatorname{div}[(v^n)^2 - 1 + o_\epsilon] \nabla u^n + I, \quad (4.29)$$

and

$$\frac{v^{n+1} - v^n}{\Delta t} - \epsilon \Delta v^{n+1} + \frac{1}{4\epsilon} v^{n+1} = -\mu |\nabla u^n|^2 v^n + \frac{1}{4\epsilon}. \quad (4.30)$$

Then, we resort to FFT to solve u^{n+1} and v^{n+1} .

Now, we present full algorithm as follows.

FFT-AT Algorithm

1. Initialization: set $o_\epsilon = \epsilon^{\hat{p}}$ ($\hat{p} > 1$), and input $u^0, v^0, \mu, \nu, \hat{p}, \epsilon, \Delta t$.
 2. For $n = 0, 1, \dots$,
 - (i) Update u^{n+1} by (4.29);
 - (ii) Update v^{n+1} by (4.30).
 3. Endfor till some stopping rule meets.
-

Now, numerical results are provided to show the efficiency and robustness of the proposed algorithms. We will compare the performance of the following four algorithms: FFT-AT algorithm as above, the Canny edge detector [29] (based on the code ‘edge(I, ‘canny’, THRESH)’ in matlab), FFT-Proximity (FFT-P) algorithm and FFT-Split Bregman (FFT-SP) algorithm. We use the relative dynamic error as in Chapter 3 (cf. (3.47)), that is,

$$E_{l^1}(\psi^{n+1}, \psi^n) := \frac{\|\psi^{n+1} - \psi^n\|_{l^1}}{\|\psi^{n+1}\|_{l^1}} \leq \eta, \quad (4.31)$$

where $\|\psi\|_{l^1} = \sum |\psi_i|$, and the summation is over all the pixels, for a prescribed tolerance $\eta > 0$ as the stopping rule, for all experiments below. For FFT-AT algorithm, the relative dynamic error is applied to the edge detector v .

In the first experiment, we test two clean images (see Figure 4.2(a) and Figure 4.2(f)), and generally choose the same parameters

$$\mu = 100, \quad \nu = 10, \quad \Delta t = 5 \times 10^{-4}, \quad r = 100,$$

for FFT-P algorithm, FFT-SP algorithm and FFT-AT algorithm, and different parameters $\lambda = 0.01$ in FFT-SP algorithm, $\epsilon = 0.01, \hat{p} = 2$ in FFT-AT algorithm and $\alpha = 0.01, \beta = 0.01$ in FFT-P algorithm, respectively. We present in Figure 4.2 the input images and the results detected by four different algorithms by using the stopping rule (4.31) with $\eta = 10^{-2}$. We observe from Figure 4.2 that FFT-P algorithm and FFT-SP algorithm outperform the Canny method and FFT-AT algorithm. The proposed two algorithms are able to detect all the meaningful edges.

In the second experiments, we turn to the comparison of four algorithms for the noisy images in Figure 4.3(a) and Figure 4.3(f) (small gaussian noise with $\sigma = 0.01$) and Figure 4.3(k) and Figure 4.3(p) (large gaussian noise with $\sigma = 0.04$), respectively. We choose the same parameters as the previous example, and present in Figure 4.3 the original images and the detected results obtained by four different algorithms with the stopping rule $\eta = 10^{-2}$. We observe from Figure 4.3(a) -4.3(j) that when the noise level is low, all algorithms produce satisfactory results. However, when we increase the level of noise, canny method fails to find the correct edges, while the other three algorithms lead to acceptable results, but FFT-P algorithm and FFT-SP algorithm yield better results and FFT-AT algorithm smoothes some details of the edge.

Figure 4.2: Comparison of four algorithms for clean images. Column 1: the original images; Column 2: edge detected by Canny method; Column 3: edge detected by FFT-AT algorithm; Column 4: edge detected by FFT-P algorithm; Column 5: edge detected by FFT-SP algorithm. From the numerical results, we can get FFT-P algorithm and FFT-SP algorithm produce better results than the other two algorithms.

4.3.2 Selection of the parameters α and β

Hereafter, we provide some insights into the selection of parameters α and β . We test FFT-P algorithm with different α and β in Figure 4.4(a) and Figure 4.4(f). We fix the other parameters as before, but choose α, β ranging from 0.01 to 0.5. In all tests, the algorithm is terminated by (4.31) with $\eta = 10^{-2}$. We see that FFT-P algorithm enjoys a similar performance in these tests.

4.3.3 More tests and extension to color images

We next test more images by using the FFT-P algorithm and FFT-SP algorithm in Figure 4.5. In these tests, both algorithms are terminated by (4.31) with $\eta = 10^{-2}$. We refer to the caption of Figure 4.5 for the choices of the parameters of these two algorithms. It is seen that both the two algorithms yield good results.

Figure 4.3: Comparison of four algorithms for noisy images. Column 1: the original images; Column 2: edge detected by Canny method; Column 3: edge detected by FFT-AT algorithm; Column 4: edge detected by FFT-P algorithm; Column 5: edge detected by FFT-SP algorithm. The iteration is stopped by (4.31) with $\eta = 10^{-2}$. FFT-P algorithm and FFT-SP algorithm produce better results than the other two algorithms.

Figure 4.4: Results obtained by FFT-P algorithm with different α and β . Row 1: results obtained with different α ; Row 2: results obtained with different β . The iteration is terminated by (4.31) with $\eta = 10^{-2}$, and we indicate the value of α and β .

Finally, we extend the proposed model and algorithm for edge detection involving color images. The color image contains three components: red, green and blue. The value of the image at a given point is a vector. Hence, we denote

$$\mathbf{I} = (I_r, I_g, I_b), \quad \mathbf{u} = (u_r, u_g, u_b), \quad \boldsymbol{\psi} = (\psi_r, \psi_g, \psi_b).$$

Meanwhile, we introduce the multichannel TV (cf. [130, 21, 113]):

$$TV(\boldsymbol{\psi}) = \int_{\Omega} \sqrt{\sum_{i=r,g,b} |\nabla \psi_i|^2} d\mathbf{x}.$$

The counterpart of the model (4.3) takes the form

$$\min_{\mathbf{u}, \boldsymbol{\psi}} \left\{ E_{CL}(\mathbf{u}, \boldsymbol{\psi}) = \mu \int_{\Omega} (1 - \boldsymbol{\psi})^2 |\nabla \mathbf{u}|^2 d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |\mathbf{u} - \mathbf{I}|^2 d\mathbf{x} + TV(\boldsymbol{\psi}) \right\}.$$

Similarly, the algorithm in Section 4.2 can be straightforwardly extended to this vectorial setting. Here, we just sketch the split Bregman method for solving $\boldsymbol{\psi}$.

We start with introducing the new variable:

$$\mathbf{d} = (\mathbf{d}_r, \mathbf{d}_g, \mathbf{d}_b) = \nabla \boldsymbol{\psi} = (\nabla \psi_r, \nabla \psi_g, \nabla \psi_b),$$

(a) Original image

(b) FFT-P

(c) FFT-SP

(d) Original image

(e) FFT-P

(f) FFT-SP

(g) Original image

(h) FFT-P

(i) FFT-SP

(j) Original image

(k) FFT-P

(l) FFT-SP

Figure 4.5: Results obtained by FFT-P algorithm and FFT-SP algorithm. In the computation, we take $\mu = 100, \mu = 10, r = 100, \lambda = 0.01, \alpha = 0.1, \beta = 0.1$. The iteration is terminated by (4.31) with $\eta = 10^{-2}$. Both FFT-P algorithm and FFT-SP algorithm can produce satisfactory results.

and apply the split Bregman iteration

$$\begin{aligned}
 (\psi^{n+1}, \mathbf{d}^{n+1}) = \arg \min_{\psi, \mathbf{d}} & \left\{ \int_{\Omega} \sqrt{\sum_{i=r,g,b} |\mathbf{d}_i|^2} d\mathbf{x} + \mu \int_{\Omega} (1 - \psi)^2 |\nabla \mathbf{u}|^2 d\mathbf{x} \right. \\
 & \left. + \frac{\lambda}{2} \sum_{i=r,g,b} \int_{\Omega} |\mathbf{d}_i - \nabla \psi_i - \mathbf{b}_i^n|^2 d\mathbf{x} + \frac{r}{2} \int_{\Omega} (\mathbf{q} - \psi)^2 d\mathbf{x} \right\},
 \end{aligned}$$

and

$$\mathbf{b}^{n+1} = \mathbf{b}^n + (\nabla \boldsymbol{\psi}^{n+1} - \mathbf{d}^{n+1}).$$

Thus, we can solve ψ_i^{n+1} as (4.26) with fixed $\mathbf{d}$ for each $i = r, g, b$, respectively. Meanwhile, the solution of $\mathbf{d}^{n+1}$ with fixed $\boldsymbol{\psi}$ can be obtained by the multi-dimensional shrinkage formula (cf. [138]):

$$\mathbf{d} = \max \left\{ \mathbf{h} - \frac{1}{\lambda}, 0 \right\} \frac{\mathbf{h}_1}{\mathbf{h}},$$

where

$$\mathbf{h}_1 = \nabla \boldsymbol{\psi} + \mathbf{b}^n, \quad \mathbf{h} = |\nabla \boldsymbol{\psi} + \mathbf{b}^n|.$$

We test the algorithms on the color images with Gaussian noise ($\sigma = 0.01$), and refer to the caption of Figure 4.6 for the choices of parameters.

Indeed, there has been limited work on edge detection involving color images. Our proposed model and algorithms provide viable means for such a purpose.

(a) Original image

(b) Noisy image

(c) FFT-SP

(d) Original image

(e) Noisy image

(f) FFT-SP

Figure 4.6: Results obtained by FFT-SP algorithm with the parameters: $\mu = 1000$, $\mu = 100$, $r = 100$, $\lambda = 1e - 2$. The iteration is stopped by (4.31) with $\eta = 10^{-2}$.

A Decomposition Model for Piecewise Smooth Image Segmentation

While intensive research has been devoted to piecewise constant image segmentation, limited works are available for segmenting images with piecewise smooth structures. Compared with the piecewise constant case, one significant difficulty lies in how to efficiently handle the coupling of the smooth approximation u of the given image I , and the edge set Γ in the full Mumford-Shah model (1.1). The remarkable attempt by Vese and Chan [135] essentially employs the level-set method in [98] to reformulate the Mumford-Shah model, and this requires to solve the approximation $\{u_i\}$ (satisfying elliptic-type equations) within the subdomain of $\{\Omega_i\}$ (identified by the level-set functions) in the evolution of the level-set functions. Indeed, with a good initialization, the proposed approach produces satisfactory segmentation results. However, in practice, one would expect to alleviate the dependence of u on Γ , for example, just to resolve u on the whole domain Ω .

In modeling images with intensity inhomogeneity [103, 141], the following model was proposed:

$$I(x) = b(x)c_i + n(\mathbf{x}), \quad \mathbf{x} \in \Omega_i, \quad 1 \leq i \leq m,$$

where $\{\Omega_i\}_{i=1}^m$ is a partition of Ω as before, b is a smooth function that models the bias field (i.e., intensity inhomogeneity) in magnetic resonance (MR) images, $\{c_i\}$ are the constant intensity values, and n models the noise. One important assumption in this context is that the bias field b is slowly varying and to take value around 1. In other words, the image of interest consists of piecewise constant structure with a “small perturbation” induced by the bias field. To search for the bias field b and $\{c_i\}$, one may resort to minimizing the functional originated from the Mumford-Shah model (1.1):

$$E(b, \mathbf{c}, \Gamma) = \mu \int_{\Omega} |\nabla b|^2 d\mathbf{x} + \frac{\nu}{2} \sum_{i=1}^m \int_{\Omega_i} |I - bc_i|^2 d\mathbf{x} + |\Gamma|, \quad (5.1)$$

(see, e.g., [69]), where $\mathbf{c} = (c_1, \dots, c_m)$. We see from the first term that the smooth field b will be solved on Ω .

In this chapter, we propose a decomposition model:

$$I(\mathbf{x}) = \lambda b(\mathbf{x}) + (1 - \lambda) \sum_{i=1}^m c_i \chi_i(\mathbf{x}) + n(\mathbf{x}), \quad \text{a.e. in } \Omega, \quad (5.2)$$

where $\lambda > 0$ is a constant, and $\{\chi_i\}$ are the characteristic functions of $\{\Omega_i\}$. We basically view a clear image (i.e., $I - n$) as a combination of a smooth field and piecewise constant structures. This can be thought of as a texture and cartoon decomposition in image processing (see, e.g., [11]). It is worthwhile to point out that image decomposition with different norms to model different components has subjected to an extensive study in e.g., [87, 57, 10] and the references therein. In this context, we plan to integrate the model (5.2) with the Mumford-Shah model with an aim to coming up a robust model for multiphase piecewise smooth image segmentation. We also focus on the efficient algorithms for its minimization and show its advantages over the existing models and approaches.

This chapter is organized as follows. In Section 5.1, we introduce the model resulted from an integration of the decomposition model with the Mumford-Shah model. We present in Section 5.2 the relaxation technique and some analysis of the model. In Section 5.3, we introduce the algorithms to solve the minimization

problem, which consist of the Fourier method, Proximity algorithm and Smoothed Chambolle-Dual algorithm. Some relevant methods for comparison and numerical results are presented in Section 5.4.

5.1 The Decomposition Model

We begin with introducing the model resulted from an integration of the decomposition (5.2) with the Mumford-Shah model (1.1). The important issue is to use appropriate norms in the energy functional to model the smooth and piecewise constant components.

As before, let $\{v_i\}_{i=1}^m$ be the characteristics of $\{\Omega_i\}_{i=1}^m$ (a nonoverlapping partition of Ω , see (1.38)-(1.41)). More precisely, we adopt H^1 -regularization for the smooth component and total variation regularization for the length of the edges. This leads to the energy functional:

$$\begin{aligned} E(b, \mathbf{c}, \mathbf{v}) := & \sum_{i=1}^m \left(\mu \lambda^2 \int_{\Omega} |\nabla b|^2 v_i \, d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |I - \lambda b - (1 - \lambda)c_i|^2 v_i \, d\mathbf{x} \right. \\ & \left. + \int_{\Omega} |\nabla v_i| \, d\mathbf{x} \right), \quad \mu, \nu, \lambda > 0, \end{aligned} \quad (5.3)$$

where $\mathbf{v} = (v_1, \dots, v_m)$ and $\mathbf{c} = (c_1, \dots, c_m)$. At this point, it is important to point out that $\mathbf{v} \in \mathcal{A}$ (defined in (1.9)), that is,

$$v_i \in \{0, 1\}, \quad 1 \leq i \leq m; \quad \sum_{i=1}^m v_i \equiv 1.$$

As $\{c_i\}$ characterize the piecewise constant structures, we would expect they return the mean of each phase, defined by

$$c_i = \frac{\int_{\Omega} I v_i \, d\mathbf{x}}{\int_{\Omega} v_i \, d\mathbf{x}}, \quad i = 1, \dots, m. \quad (5.4)$$

It is also for convenience of implementation that we search for the triple $(b, \mathbf{c}, \mathbf{v})$

by minimizing the functional:

$$\begin{aligned}
 E_{PS}(b, \mathbf{c}, \mathbf{v}) := & \mu \lambda^2 \sum_{i=1}^m \int_{\Omega} |\nabla b|^2 v_i d\mathbf{x} + \nu \lambda^2 \int_{\Omega} |I - b|^2 d\mathbf{x} \\
 & + \sum_{i=1}^m \left((1 - \lambda)^2 \nu \int_{\Omega} |I - c_i|^2 v_i d\mathbf{x} + \int_{\Omega} |\nabla v_i| d\mathbf{x} \right).
 \end{aligned} \tag{5.5}$$

Noticing that $I = \lambda I + (1 - \lambda)I$, we infer that

$$E(b, \mathbf{c}, \mathbf{v}) \leq E_{PS}(b, \mathbf{c}, \mathbf{v}).$$

With this setup, (5.4) can be achieved, and it appears more accessible for implementation to be shown shortly.

5.2 Relaxation and Well-posedness

We reiterate that one obvious challenge for minimizing (5.5) is that $\{v_i\}$ must take the binary values, so the admissible set $\mathcal{A}$ is non-convex. Although the direct approach presented in Chapter 3 could be explored in this setting, we would like to relax $\mathcal{A}$ into the convex set

$$\Delta_+ := \left\{ \mathbf{v} := (v_1, \dots, v_m) \in (BV(\Omega))^m : v_i \in [0, 1]; \sum_{i=1}^m v_i = 1 \right\}, \tag{5.6}$$

where $BV(\Omega)$ is the space of bounded variation as defined (1.6). Such a relaxation is commonly-used in literature (see, e.g., [144, 64, 14, 69]). Letting $\{v_i\}$ vary in $[0, 1]$, instead of only taking the binary value 0 or 1, is preferred in this context, (at least) as this allows for certain fuzzyness for the classification of pixels near the discontinuities, and this favors the global smooth of b .

Next, we show the existence of the minimizer of the minimization problem

$$\min_{(b, \mathbf{c}, \mathbf{v}) \in \Lambda_m} E_{PS}(b, \mathbf{c}, \mathbf{v}), \tag{5.7}$$

where E_{PS} is defined in (5.5) and the solution space is

$$\Lambda_m = \{(b, \mathbf{c}, \mathbf{v}) : b \in H^1(\Omega); \mathbf{v} \in \Delta_+; \mathbf{c} \text{ is constant-valued vector}\}. \tag{5.8}$$

Here, we follow the argument in [69, 117].

Theorem 5.1. *Suppose that the image $I \in L^\infty(\Omega)$ and $I \geq 0$. Then for fixed μ, ν and λ , there exists a minimizer $(b^*, \mathbf{c}^*, \mathbf{v}^*) \in \Lambda_m$ of (5.7)-(5.8).*

Proof. For any $(b, \mathbf{c}, \mathbf{v}) \in \Lambda_m$ and $I \in L^\infty(\Omega)$, we have $E_{PS}(b, \mathbf{c}, \mathbf{v}) < \infty$. Let $\{(b^n, \mathbf{c}^n, \mathbf{v}^n)\} \in \Lambda_m$ be a minimizing sequence such that

$$E_{PS}(b^n, \mathbf{c}^n, \mathbf{v}^n) \rightarrow \min E_{PS}(b, \mathbf{c}, \mathbf{v}) \quad \text{as } n \rightarrow \infty. \quad (5.9)$$

Hereafter, let C be a generic positive constant independent of n . We infer from (5.9) that

$$\begin{aligned} E_{PS}(b^n, \mathbf{c}^n, \mathbf{v}^n) &= \mu\lambda^2 \sum_{i=1}^m \int_{\Omega} |\nabla b^n|^2 v_i^n d\mathbf{x} + \nu\lambda^2 \int_{\Omega} |I - b^n|^2 d\mathbf{x} \\ &\quad + (1-\lambda)^2 \nu \sum_{i=1}^m \int_{\Omega} |I - c_i^n|^2 v_i^n d\mathbf{x} + \sum_{i=1}^m \int_{\Omega} |\nabla v_i^n| d\mathbf{x} \\ &\leq C, \end{aligned}$$

which implies

$$\mu\lambda^2 \sum_{i=1}^m \int_{\Omega} |\nabla b^n|^2 v_i^n d\mathbf{x} \leq C. \quad (5.10)$$

$$\nu\lambda^2 \int_{\Omega} |I - b^n|^2 d\mathbf{x} \leq C. \quad (5.11)$$

$$\int_{\Omega} |\nabla v_i^n| d\mathbf{x} \leq C. \quad (5.12)$$

We obtain from (5.11) that

$$\begin{aligned} \int_{\Omega} |b^n|^2 d\mathbf{x} &= \int_{\Omega} |I - b^n - I|^2 d\mathbf{x} \\ &\leq 2 \int_{\Omega} |I - b^n|^2 d\mathbf{x} + 2 \int_{\Omega} |I|^2 d\mathbf{x} \\ &\leq C. \end{aligned}$$

As $\mathbf{v} \in \Delta_+$, we have

$$\int_{\Omega} |\nabla b^n|^2 d\mathbf{x} = \sum_{i=1}^m \int_{\Omega} |\nabla b^n|^2 v_i d\mathbf{x} \leq C.$$

Thus, we conclude from the above two estimates that

$$\|b^n\|_{H^1(\Omega)} = \|b^n\|_{L^2(\Omega)} + \|\nabla b^n\|_{L^2(\Omega)} \leq C.$$

By the L^2 -weak compactness of bounded H^1 -sequence (cf. [1]), there is a subsequence, still denoted by $\{b^n\}$ and b^* , which satisfies

$$b^n \rightharpoonup b^* \text{ weakly in } H^1(\Omega).$$

By the compact Sobolev imbedding theorem (cf. [1]), we can extract a subsequence from $\{b^n\}$, also denoted by $\{b^n\}$, such that

$$b^n \rightarrow b^* \text{ strongly in } L^2(\Omega); \quad (5.13a)$$

$$b^n \rightarrow b^* \text{ a.e. in } \Omega; \quad (5.13b)$$

$$b^n \rightharpoonup b^* \text{ weakly in } H^1(\Omega). \quad (5.13c)$$

Hence, by the lower semicontinuity,

$$\int_{\Omega} |I - b^*|^2 d\mathbf{x} \leq \liminf_{n \rightarrow \infty} \int_{\Omega} |I - b^n|^2 d\mathbf{x}. \quad (5.14)$$

By the compactness of sequence $\{c_i^n\}$ and the Arzela-Asoli compactness criterion (cf. [8, 9]), there exists a subsequence, also denoted by $\{c_i^n\}$ and $\{c_i^*\}$, such that

$$c_i^n \rightarrow c_i^* \text{ uniformly, } 1 \leq i \leq m.$$

Moreover, we obtain from the fact $\mathbf{v} \in \Delta_+$ that

$$\|v_i^n\|_{L^1(\Omega)} = \int_{\Omega} |v_i^n| d\mathbf{x} \leq |\Omega|, \quad 1 \leq i \leq m,$$

where $|\Omega|$ is the area of Ω . Combined with (5.12), it follows that $\{v_i^n\}$ is uniformly bounded in $BV(\Omega)$. By the compactness property of BV space (cf. [47]), for $1 \leq i \leq m$, there exists a subsequence, also denoted by $\{v_i^n\}$ and $\{v_i^*\}$, such that

$$v_i^n \rightarrow v_i^* \text{ strongly in } L^1(\Omega); \quad (5.15a)$$

$$v_i^n \rightarrow v_i^* \text{ a.e. in } \Omega; \quad (5.15b)$$

$$\nabla v_i^n \rightarrow \nabla v_i^* \text{ in the sense of measure.} \quad (5.15c)$$

Using the lower semicontinuity of the total variation (cf. [47]), we obtain

$$\int_{\Omega} |\nabla v_i^*| d\mathbf{x} \leq \liminf_{n \rightarrow \infty} \int_{\Omega} |\nabla v_i^n| d\mathbf{x}. \quad (5.16)$$

By the convergence result, we have

$$v_i^* \in [0, 1] \quad \text{and} \quad \sum_{i=1}^m v_i^* = 1.$$

Finally, it is apparent that

$$b^n \rightarrow b^*, \quad v_i^n \rightarrow v_i^* \quad \text{a.e. in } \Omega, \quad \text{and} \quad c_i^n \rightarrow c_i^*.$$

Using the Fatou's Lemma (cf. [111]), we get

$$\begin{aligned} & \mu\lambda^2 \int_{\Omega} |\nabla b^*|^2 v_i^* d\mathbf{x} + (1-\lambda)^2 \nu \int_{\Omega} |I - c_i^*|^2 v_i^* d\mathbf{x} \\ & \leq \liminf_{n \rightarrow \infty} \mu\lambda^2 \int_{\Omega} |\nabla b^n|^2 v_i^n d\mathbf{x} + (1-\lambda)^2 \nu \int_{\Omega} |I - c_i^n|^2 v_i^n d\mathbf{x}. \end{aligned} \quad (5.17)$$

Combining the inequalities (5.14), (5.16) and (5.17), we arrive at

$$E_{PS}(b^*, \mathbf{c}^*, \mathbf{v}^*) \leq \liminf_{n \rightarrow \infty} E_{PS}(b^n, \mathbf{c}^n, \mathbf{v}^n) = \min E_{PS}(b, \mathbf{c}, \mathbf{v}).$$

Thus, $(b^*, \mathbf{c}^*, \mathbf{v}^*)$ is a minimizer of (5.7)-(5.8). $\square$

5.3 Minimization Algorithm

We adopt the splitting technique and seek the minimizer $(b, \mathbf{c}, \mathbf{v})$ separately as follows.

(a). Fixing $\mathbf{v}$, we search for b as a solution of

$$\min_{b \in H^1(\Omega)} \left\{ \mu\lambda^2 \sum_{i=1}^m \int_{\Omega} |\nabla b|^2 v_i d\mathbf{x} + \nu\lambda^2 \int_{\Omega} |I - b|^2 d\mathbf{x} \right\}. \quad (5.18)$$

(b). Fixing $\mathbf{v}$, we search for $\mathbf{c}$ as a solution of

$$\min_{\mathbf{c}} \left\{ (1-\lambda)^2 \nu \sum_{i=1}^m \int_{\Omega} |I - c_i|^2 v_i d\mathbf{x} \right\}.$$

(c). Fixing b and $\mathbf{c}$, we search for $\mathbf{v}$ as a solution of

$$\min_{\mathbf{v} \in \Delta_+} \left\{ \mu \lambda^2 \sum_{i=1}^m \int_{\Omega} |\nabla b|^2 v_i d\mathbf{x} + (1 - \lambda)^2 \nu \sum_{i=1}^m \int_{\Omega} |I - c_i|^2 v_i d\mathbf{x} + \sum_{i=1}^m \int_{\Omega} |\nabla v_i| d\mathbf{x} \right\}. \quad (5.19)$$

As expected, the use of (5.5) (rather than (5.3)) leads to the expression of $\mathbf{c}$ in (5.4). Next, we present the algorithms for (5.18) and (5.19).

5.3.1 Minimization with respect to b

Like in Subsection 4.2.1, the Fourier method can be used for finding b in (5.18). In this situation, the corresponding Euler-Lagrange equation takes the form

$$\begin{cases} -\mu \sum_{i=1}^m \operatorname{div}(\nabla b v_i) + \nu(b - I) = 0, \\ \frac{\partial b}{\partial \mathbf{n}} \Big|_{\partial \Omega} = 0. \end{cases}$$

The counterpart of (4.11) is

$$\begin{cases} -\mu \Delta b^{k+1} + \nu b^{k+1} = W(b^k), & \text{in } \Omega, \\ \frac{\partial b^{k+1}}{\partial \mathbf{n}} \Big|_{\partial \Omega} = 0, & k = 0, 1, \dots, \end{cases} \quad (5.20)$$

where

$$W(b) = \mu \sum_{i=1}^m \operatorname{div}(\nabla b v_i) - \mu \Delta b + \nu I.$$

Then, the equation (5.20) can be solved efficiently by using the fast Fourier transform (FFT). One may refer to [118] for the implementation of FFT in Matlab.

5.3.2 Smoothed Chambolle-Dual algorithm for solving $\mathbf{v}$

Now, we focus on the subproblem (5.19). As described in Section 5.2, it is desirable to construct algorithm over the relaxed admissible set Δ_+ . Hereafter, we

handle the constraints by using the entropy penalization (cf. [14]), which results in convex subproblems of $\{v_i\}$. Using this notion, we convert (5.19) into

$$\begin{aligned} \min_v \Big\{ & \mu\lambda^2 \sum_{i=1}^m \int_{\Omega} |\nabla b|^2 v_i d\mathbf{x} + (1-\lambda)^2 \nu \sum_{i=1}^m \int_{\Omega} |I - c_i|^2 v_i d\mathbf{x} \\ & + \sum_{i=1}^m \int_{\Omega} |\nabla v_i| d\mathbf{x} + r \sum_{i=1}^m \int_{\Omega} v_i \log v_i d\mathbf{x} \Big\}, \quad \mu, \nu, \lambda, r > 0. \end{aligned} \quad (5.21)$$

Defining

$$f_i = \mu\lambda^2 |\nabla b|^2 + (1-\lambda)^2 \nu |I - c_i|^2, \quad (5.22)$$

we derive from the definition of the total variation (see (1.4)-(1.5)) that

$$\min_v \max_{\vec{\mathbf{p}} \in \mathbf{S}} \left\{ \sum_{i=1}^m \int_{\Omega} (v_i f_i + r v_i \log v_i) d\mathbf{x} + \sum_{i=1}^m \int_{\Omega} v_i \operatorname{div} \mathbf{p}_i d\mathbf{x} \right\}, \quad (5.23)$$

where we recall that $\vec{\mathbf{p}} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m)$, and $\mathbf{S} = S^m$ (where S is defined by (1.5) in Chapter 1).

Next, we apply the Chambolle's dual algorithm [34] to (5.23). The starting point is to exchange “min” and “max”, as it satisfies the condition of the minimax theorem (cf. [51, 53]). This leads to

$$\max_{\vec{\mathbf{p}} \in \mathbf{S}} \min_v \left\{ \sum_{i=1}^m \int_{\Omega} (v_i f_i + r v_i \log v_i) d\mathbf{x} + \sum_{i=1}^m \int_{\Omega} v_i \operatorname{div} \mathbf{p}_i d\mathbf{x} \right\}. \quad (5.24)$$

For fixed $\vec{\mathbf{p}}$, the optimality condition of the “min” problem gives

$$v_i = \exp\left(-\frac{\operatorname{div} \mathbf{p}_i + f_i + r}{r}\right), \quad 1 \leq i \leq m. \quad (5.25)$$

In practice, it is necessary to meet the constraint $\sum_{i=1}^m v_i = 1$ (see, e.g., [143] for such a treatment), so we compute

$$v_i = \frac{\exp\left(-\frac{\operatorname{div} \mathbf{p}_i + f_i}{r}\right)}{\sum_{j=1}^m \exp\left(-\frac{\operatorname{div} \mathbf{p}_j + f_j}{r}\right)} = \frac{1}{\sum_{j=1}^m \exp\left(-\frac{(\operatorname{div} \mathbf{p}_j + f_j) - (\operatorname{div} \mathbf{p}_i + f_i)}{r}\right)}, \quad (5.26)$$

for $1 \leq i \leq m$.

Plugging (5.26) into (5.23) leads to the dual model:

$$\min_{\vec{p} \in \mathcal{S}} \int_{\Omega} \left\{ \log \sum_{i=1}^m \exp \left(\frac{-f_i - \operatorname{div} \mathbf{p}_i}{r} \right) \right\} d\mathbf{x}. \quad (5.27)$$

Following the derivation of the Chambolle's dual algorithm [34] (also see Subsection 3.1.1), we obtain the nonlinear equation:

$$-\nabla \left(\frac{\exp \left(-\frac{\operatorname{div} \mathbf{p}_i + f_i}{r} \right)}{\sum_{j=1}^m \exp \left(-\frac{\operatorname{div} \mathbf{p}_j + f_j}{r} \right)} \right) + \left| \nabla \left(\frac{\exp \left(-\frac{\operatorname{div} \mathbf{p}_i + f_i}{r} \right)}{\sum_{j=1}^m \exp \left(-\frac{\operatorname{div} \mathbf{p}_j + f_j}{r} \right)} \right) \right| \mathbf{p}_i = 0,$$

for $1 \leq i \leq m$. In view of (5.26), we write it the compact form:

$$-\nabla v_i + |\nabla v_i| \mathbf{p}_i = 0, \quad 1 \leq i \leq m. \quad (5.28)$$

It can be solved by a semi-implicit discretization in time as in [34]. Let τ be the time step size and $\vec{p}^n$ be the approximation of $\vec{p}$ at $n\tau$. The scheme reads

$$\frac{\mathbf{p}_i^{n+1} - \mathbf{p}_i^n}{\tau} = \nabla v_i^n - |\nabla v_i^n| \mathbf{p}_i^{n+1} \Rightarrow \mathbf{p}_i^{n+1} = \frac{\mathbf{p}_i^n + \tau \nabla v_i^n}{1 + \tau |\nabla v_i^n|}, \quad 1 \leq i \leq m. \quad (5.29)$$

We summarize the full algorithm for the model (5.7)-(5.8), termed as the FFT-Smoothed-Chamboll-Dual (FFT-SCD) Algorithm.

FFT-SCD Algorithm

1. Initialization: set $\vec{p}^0 = \mathbf{0}$ and input $\mathbf{v}^0, b^0, r, \mu, \nu, \lambda, \tau$.
2. For $n = 0, 1, \dots$
 - (i) Compute $\mathbf{c}^n$ by (5.4) with $\mathbf{v}^n$ in place of $\mathbf{v}$;
 - (ii) Update b^{n+1} using FFT and the iteration scheme (5.20) with initial value b^n and $\mathbf{v}^n$ in place of $\mathbf{v}$;
 - (iii) Compute $\mathbf{v}^{n+1}$ by

$$v_i^{n+1} = \frac{\exp \left(-\frac{\operatorname{div} \mathbf{p}_i^n + f_i^n}{r} \right)}{\sum_{j=1}^m \exp \left(-\frac{\operatorname{div} \mathbf{p}_j^n + f_j^n}{r} \right)}, \quad 1 \leq i \leq m,$$

where

$$f_i^n = \mu \lambda^2 |\nabla b^{n+1}|^2 + (1 - \lambda)^2 \nu |I - \mathbf{c}_i^n|^2;$$

(iv) Compute $\bar{\mathbf{p}}^{n+1}$ by

$$\mathbf{p}_i^{n+1} = \frac{\mathbf{p}_i^n + \tau \nabla v_i^{n+1}}{1 + \tau |\nabla v_i^{n+1}|}, \quad 1 \leq i \leq m.$$

3. Endfor till some stopping rule meets.

Different from the dual algorithm presented in Chapter 3, we adopt the relaxation of admissible set and implement the entropy penalization for the “relaxed” constraints. This allows for certain fuzziness and efficient algorithms for convex optimization. It also differs from the global smoothed-dual algorithm for piecewise constant case in [14] (see Chapter 3 for a brief description) in terms of the way to introduce the algorithm and compute the dual variables.

5.3.3 Proximity algorithm for solving $\mathbf{v}$

With the convex relaxation and entropy penalization, it is interesting to apply the proximity algorithm to this more complicated setting. Indeed, as we will see below, it provides a viable alternative to the preceding dual algorithm.

As described in Subsection 4.2.2, we first rewrite (5.21) as

$$\min_{\mathbf{v}} \left\{ \sum_{i=1}^m \rho_i(v_i) + \sum_{i=1}^m (g \circ B)(v_i) \right\},$$

where

$$\rho_i(v_i) = \mu \lambda^2 \int_{\Omega} |\nabla b|^2 v_i d\mathbf{x} + (1 - \lambda)^2 \nu \int_{\Omega} |I - c_i|^2 v_i d\mathbf{x} + r \int_{\Omega} v_i \log v_i d\mathbf{x},$$

and g, B are defined in (4.14).

By Lemma 4.2, we solve $\mathbf{v}$ from (4.15) and obtain

$$v_i = \text{prox}_{\frac{1}{\alpha} \rho_i} \left(v_i - \frac{\beta}{\alpha} B^* \mathbf{s}_i \right), \quad 1 \leq i \leq m, \quad (5.30)$$

and

$$\mathbf{s}_i = (\mathbb{I} - \text{prox}_{\frac{1}{\beta} g})(B v_i + \mathbf{s}_i), \quad 1 \leq i \leq m. \quad (5.31)$$

To this end, we denote $\vec{s} = (s_1, \dots, s_m)$.

To solve (5.30), we use Definition 4.2 and find

$$\text{prox}_{\frac{1}{\alpha}\rho_i}(v_i - \frac{\beta}{\alpha}B^*s_i) = \arg \min_{\omega_i} \left\{ \frac{1}{\alpha}\rho_i(\omega_i) + \frac{1}{2} \left\| \omega_i - (v_i - \frac{\beta}{\alpha}B^*s_i) \right\|_2^2 \right\},$$

for $1 \leq i \leq m$. The associated Euler-Lagrangian equation is

$$\frac{\alpha\omega_i}{r} + \log \omega_i = M_i, \quad 1 \leq i \leq m, \quad (5.32)$$

where

$$M_i = \frac{\alpha(v_i - \frac{\beta}{\alpha}B^*s_i)}{r} - \frac{\nu(1-\lambda)^2|I - c_i|^2 + \mu\lambda^2|\nabla b|^2}{r} - 1.$$

Taking exponentials on both sides, leads to

$$\omega_i e^{\frac{\alpha\omega_i}{r}} = e^{M_i}, \quad 1 \leq i \leq m. \quad (5.33)$$

Remark 5.1. To find ω_i from the above algebraic equation, we resort to the Newton's method:

$$\omega_i^{k+1} = \omega_i^k - \frac{F(\omega_i^k)}{F'(\omega_i^k)}, \quad 1 \leq i \leq m, \quad k = 0, 1, 2, \dots, \quad (5.34)$$

where $F(\omega_i) = \omega_i e^{\frac{\alpha\omega_i}{r}} - e^{M_i}$. □

Letting $\{v_i = \omega_i\}$ and imposing the constraint $\sum_{i=1}^m v_i = 1$, we obtain

$$v_i = \frac{\omega_i}{\sum_{j=1}^m \omega_j}, \quad 1 \leq i \leq m. \quad (5.35)$$

Finally, we have the closed-form formula for $\vec{s}$ in (5.31) (cf. (4.22)):

$$s_i = (Bv_i + s_i) - \max\{|Bv_i + s_i| - 1/\beta, 0\} \cdot \text{sign}(Bv_i + s_i), \quad 1 \leq i \leq m. \quad (5.36)$$

Now, we are ready to present the algorithm. We replace (iii) and (iv) in FFT-SCD algorithm by the proximity algorithm, and summarize below the full algorithm, termed as FFT-Proximity-Segmentation (FFT-PS) Algorithm.

FFT-PS Algorithm

1. Initialization: set $\vec{s}^0 = \mathbf{0}$ and input $\mathbf{v}^0, b^0, \mu, \nu, \alpha, \beta, \lambda, r$.

2. For $n = 0, 1, \dots$

- (i) Compute $\mathbf{c}^n$ by (5.4) with $\mathbf{v}^n$ in place of $\mathbf{v}$;
- (ii) Update b^{n+1} using FFT and the iteration scheme (5.20) with initial value b^n and $\mathbf{v}^n$ in place of $\mathbf{v}$;
- (iii) Update $\mathbf{v}^{n+1}$ using the iteration scheme (5.34) with initial value $\mathbf{v}^n$ with $b^{n+1}, \mathbf{c}^n, \bar{\mathbf{s}}^n$ in place of $b, \mathbf{c}, \bar{\mathbf{s}}$, respectively;
- (iv) Compute $\bar{\mathbf{s}}^{n+1}$ by

$$\mathbf{s}_i^{n+1} = (Bv_i^{n+1} + \mathbf{s}_i^n) - \max\{|Bv_i^{n+1} + \mathbf{s}_i^n| - 1/\beta, 0\} \cdot \text{sign}(Bv_i^{n+1} + \mathbf{s}_i^n),$$

for $1 \leq i \leq m$.

3. Endfor till some stopping rule meets.

Compared with the FFT-SCD algorithm, special care has to be taken for solving (5.33), though there is no need to solve differential equations (for the dual variables). Note that $B = \nabla$ and $B^* = -\text{div}$ as before.

5.4 Numerical Results

In this section, we provide ample numerical experiments to demonstrate the performance of the decomposition model and proposed algorithms. We first test them on the decomposition of noisy signals into piecewise constant components (obtained from $\{v_i\}$) and smooth component (contained in b), as in (5.2). Then we focus on the comparison with the relevant approaches in Vese-Chan [135].

As before, we use the relative dynamic error as the stopping rule:

$$E_{l^1}(\mathbf{v}^{n+1}, \mathbf{v}^n) := \frac{\|\mathbf{v}^{n+1} - \mathbf{v}^n\|_{l^1}}{\|\mathbf{v}^{n+1}\|_{l^1}} \leq \eta, \quad (5.37)$$

for a prescribed tolerance $\eta > 0$.

5.4.1 One-dimensional tests

For clarity and ease of understanding our approach, we particularly test it on the signal with piecewise constant and smooth structures, polluted by noise. We highlight in Figure 5.1 the components: $\sum c_i v_i$ (in pink), b (in black) and $u = \lambda b + (1 - \lambda) \sum c_i v_i$ (in red), resulted from our model and algorithms. It is seen that even for large noise, very accurate segmentation can be obtained. Relatively, the FFT-SCD algorithm is robust for larger noise. This will be also verified by the two-dimensional tests below.

5.4.2 The Vese-Chan method [135]

Here, we intend to compare our approach with the Vese-Chan method [135] for piecewise smooth images. For clarity, we mainly discuss the two-phase framework. Let ϕ be signed distance function defined in (1.14), and $H(\phi)$ be the heaviside function in (1.15). Then two-phase model reads

$$\begin{aligned} \min_{u^+, u^-, \phi} & \left\{ \frac{\nu}{2} \int_{\Omega} |u^+ - I|^2 H(\phi) d\mathbf{x} + \frac{\nu}{2} \int_{\Omega} |u^- - I|^2 (1 - H(\phi)) d\mathbf{x} \right. \\ & \left. + \mu \int_{\Omega} |\nabla u^+|^2 H(\phi) d\mathbf{x} + \mu \int_{\Omega} |\nabla u^-|^2 (1 - H(\phi)) d\mathbf{x} + \int_{\Omega} |\nabla H(\phi)| d\mathbf{x} \right\}, \end{aligned} \quad (5.38)$$

where the smooth approximation u is split into two parts:

$$u(\mathbf{x}) = \begin{cases} u^+(\mathbf{x}), & \text{if } \phi(\mathbf{x}) \geq 0, \\ u^-(\mathbf{x}), & \text{if } \phi(\mathbf{x}) < 0. \end{cases}$$

In real implementation, the smear-out Heaviside function H_ϵ and delta function δ_ϵ , defined in (1.21), are used to replace H and δ . Then, The optimality conditions of (5.38) lead to

$$\begin{aligned} -\frac{2\mu}{\nu} \Delta u^+ + u^+ &= I, \quad \text{in } \{\phi(\mathbf{x}) \geq 0\}, \\ -\frac{2\mu}{\nu} \Delta u^- + u^- &= I, \quad \text{in } \{\phi(\mathbf{x}) < 0\}, \end{aligned}$$

Figure 5.1: Results obtained by FFT-SCD algorithm and FFT-PS algorithm for one-dimensional tests with $\mu = 0.1, \nu = 0.01, \alpha = 0.01, \beta = 0.01, r = 0.1, \lambda = 0.8, \tau = 0.01$. Column 1: the clean images; Column 2: the noisy images; Column 3-4: the segmented results u together with the smoothed part b and the jumps $\sum c_i v_i$ of FFT-SCD algorithm and FFT-PS algorithm, respectively. The iteration is stopped by (5.37) with $\eta = 10^{-2}$. FFT-SCD algorithm outperform FFT-PS algorithm for the multiphase test with large noise.

and

$$\delta_\epsilon(\phi) \left[\nabla \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \frac{\nu}{2} |u^+ - I|^2 - \mu |\nabla u^+|^2 + \frac{\nu}{2} |u^- - I|^2 + \mu |\nabla u^-|^2 \right] = 0, \quad (5.39)$$

with the homogeneous Neumann boundary conditions as before.

We use the gradient descent method to solve (5.39) for fixed u^+, u^-

$$\begin{aligned} \frac{\phi^{n+1} - \phi^n}{\Delta t} = & \delta_\varepsilon(\phi^n) \left[\nabla \left(\frac{\nabla \phi^n}{|\nabla \phi^n|_\theta} \right) - \frac{\nu}{2} |u^+ - I|^2 - \mu |\nabla u^+|^2 \right. \\ & \left. + \frac{\nu}{2} |u^- - I|^2 + \mu |\nabla u^-|^2 \right], \end{aligned} \quad (5.40)$$

where $|\nabla \phi^n|_\theta = \sqrt{(\phi_x^n)^2 + (\phi_y^n)^2 + \theta}$ as before, for some small $\theta > 0$.

We solve the extension of u^+ and u^- to $\{\phi \leq 0\}$ and $\{\phi \geq 0\}$, respectively as in [135] (see the bottom-right of Page 291).

For comparison propose, we present the Vese-Cham method for multiphase piecewise smooth image in [135].

Vese-Chan method [135]

1. Initialization: input $\phi^0, u^{0,+}, u^{0,-}, \Delta t, \mu, \nu$.

2. For $n = 0, 1, \dots$

(i) Compute $u^{n+1,+}$ and $u^{n+1,-}$ by

$$\begin{aligned} -\frac{2\mu}{\nu} \Delta u^{n+1,+} + u^{n+1,+} &= I \text{ in } \{\phi^n(\mathbf{x}) \geq 0\}, \\ -\frac{2\mu}{\nu} \Delta u^{n+1,-} + u^{n+1,-} &= I \text{ in } \{\phi^n(\mathbf{x}) < 0\}, \end{aligned}$$

(ii) Compute the extension of $u^{n+1,+}$ to $\{\phi^n \leq 0\}$ and $u^{n+1,-}$ to $\{\phi^n \geq 0\}$;

(iii) Compute ϕ^{n+1} by (5.40) with $u^{n+1,+}, u^{n+1,-}$ in place of u^+, u^- , respectively.

3. Endfor till some stopping rule meets.

Now, we present in Figure 5.2 the comparison of FFT and Smoothed Chambolle-Dual (FFT-SCD) algorithm, FFT and Proximity Segmentation (FFT-PS) algorithm and the Vese-Chan method. For the Vese-Chan method, the relative dynamic error is applied to the level-set function ϕ . We test two-phase images with

different noise levels (see, Figure 5.2(a) and Figure 5.2(e)). We present in Figure 5.2 the input images and segmented results, and tabulate in Table 5.1 the iteration number and computational time obtained by three algorithms (terminated by (5.37) with $\varsigma = 10^{-2}$, and refer to the caption of Figure 5.2 for the computational setup). We find that at low noise level, all three algorithms produce satisfactory segmentation results, but FFT-SCD algorithm and FFT-PS algorithm converge much faster. However, when we increase the noise level, FFT-SCD algorithm yields better segmentation and both of FFT-SCD algorithm and FFT-PS algorithm outperform the Vese-Chan method and converge much faster (see Table 5.1).

Figure 5.2: Comparison of Vese-Chan method, FFT-SCD algorithm and FFT-PS algorithm. In the computation, we take $\mu = 0.1, \nu = 0.01, \alpha = 0.01, \beta = 0.01, r = 0.1, \lambda = 0.8, \tau = 0.01, \Delta t = 10^{-4}$. The iteration is terminated by (5.37) with $\eta = 10^{-2}$. Both of FFT-SCD algorithm and FFT-PS algorithm outperform the Vese-Chan method.

Table 5.1: Comparison of three different algorithms.

Figure 5.2(a)	iterations	time (second)
Vese-Chan method	282	78.54
FFT-SCD	86	3.45
FFT-PS	163	5.34
Figure 5.2(e)	iterations	time (second)
Vese-Chan method	512	193.76
FFT-SCD	185	6.43
FFT-PS	412	16.36

5.4.3 More two-dimensional tests

Next, we present some more two-dimensional tests. Firstly, we test a brain image in Figure 5.3(a) available to the public at “<http://www.bic.mni.mcgill.ca/brainweb/>”. For the image used in this test the noise level is 7% and the nonuniformity intensity level of the RF-puls is 20% (see the details on the website above). There are three tissue classes that should be identified: cerebrospinal fluid, gray matter and white matter. Hence, we use four-phase and refer to the caption of Figure 5.3 for the selection of parameters and time step size. We present in Figure 5.3 the original image, segmentation results of FFT-SCD algorithm and FFT-PS algorithm at the iteration terminated by (5.37) with $\eta = 10^{-2}$, and the exact segmentation results. From these figures in Figure 5.3, it can be seen that both of FFT-SCD algorithm and FFT-PS algorithm can produce a similar performance as the exact results.

Finally, some more experiments are conducted in Figure 5.4. We test two-phase images (see Figure 5.4(a) and Figure 5.4(d)), and three-phase image (see Figure 5.4(j)) by using FFT-SCD algorithm and FFT-PS algorithm. We refer to the caption of Figure 5.4 for the choices of the parameters. In all tests, both of the two algorithms provide satisfactory results.

Figure 5.3: Results obtained by FFT-SCD algorithm and FFT-PS algorithm. Row 1: segmented results of FFT-SCD algorithm. Row 2: segmented results of FFT-PS algorithm. Row 3: exact segmented results. In this computation, we choose $\mu = 0.1, \nu = 0.01, \alpha = 0.01, \beta = 0.01, r = 0.1, \lambda = 0.8, \tau = 0.01, \tau_{fs} = 1, \theta = 0.1, w = 0.001$. The iteration is stopped by (5.37) with $\eta = 10^{-2}$. FFT-SCD algorithm and FFT-PS algorithm produce comparable results as the exact results, and all these phases match the exact phases well.

Figure 5.4: Results obtained by FFT-SCD algorithm and FFT-PS algorithm with $\mu = 0.1, \nu = 0.01, \alpha = 0.01, \beta = 0.01, r = 0.1, \lambda = 0.8, \tau = 0.01$. Both of FFT-PS algorithm and FFT-SCD algorithm provide good results.

Conclusion and Future Works

In this chapter, we conclude the thesis with a summary of the works done, and elaborate on some relevant topics for further investigation.

6.1 Conclusion

The thesis was devoted to fast image segmentation based on Mumford-Shah type models. We reiterate that the main contributions lie in the following four aspects.

- (i) In Chapter 2, we developed efficient dual-algorithms for the TV-Allen-Cahn model for two-phase image segmentation. The use of a splitting-penalty techniques led to a very interesting Allen-Cahn type model with “double-well” potential different from the usual one. Consequently, we could apply the Chambolle’s dual algorithm to such a model. With a further edge-featured penalty of the dual variable, we obtained a new diffusive model, so we could construct fast algorithms using operator splitting in time and FFT in space. Various numerical results showed that the dual algorithms were much faster and stabler than the primal gradient descent approach, and the new dual algorithm was at least as efficient as the Chambolle’s algorithm, and relatively accurate.

- (ii) In Chapter 3, we proposed a direct primal-dual approach for global minimization of the multiclass labeling problem and multiphase segmentation problem. Different from the existing works, the underlying analysis and algorithms were based on a binary setting, which did not require convex relaxation. Using the augmented Lagrangian technique, we were able to find the relations between the primal and dual variables that involved almost the minimum number of parameters. The proposed algorithms could be viewed as the counterparts of the Chambolle's algorithm in the context of labeling and segmentation. Indeed, the time step size could be taken as large as that of the Chambolle's algorithm, and the computational cost turned out to be almost of the same amount. Moreover, the smoothing parameter involved in the algorithm was quite flexible to choose. Various numerical results demonstrated the advantages of the methods over the existing approaches.
- (iii) In Chapter 4, we presented efficient algorithms for the modified Mumford-Shah model proposed in [137]. We used the relaxation technique and splitting-penalty method to handle the constraint of the binary level-set function. We could propose efficient algorithms, which included the Fourier method, proximity algorithm and split Bregman algorithm. Moreover, we extended this approach to edge detection of a color image. The numerical results showed the efficiency and robustness of the proposed model and algorithms.
- (iv) In Chapter 5, we introduced a new decomposition model for multiphase piecewise smooth image segmentation. By decomposing an image into a piecewise constant part, smooth component and noise part, we modeled each component appropriately and derived a new model from the MS model. We presented the relaxation technique and result on the existence of minimizer. By using the entropy penalization to handle the involved constraints of the level-set functions, we were able to obtain a convex functional so that the techniques, e.g., the proximity algorithm, could excel their capability.

We presented ample numerical results to show that our approach outperformed some relevant methods.

6.2 Future Works

Based on the research in this thesis, we explore some issues and topics that are worthy of further study.

- Much of our previous study has been focused on the gray image segmentation. The framework can be extended color image segmentation, which needs extra efforts to take care of the vectorial TV, and the stability of the algorithm.
- We shall further apply the developed methods for real image processing. For example, the decomposition model proposed in Chapter 5 is expected to be useful processing for real MRI images with intensity inhomogeneities.
- We plan to develop new ideas for piecewise smooth image segmentation. One thought is to introduce a smoothing operator to play the role of H^1 -smoothing in the original Mumford-Shah model. Below, we just sketch the process and present preliminary results.

One difficulty in solving the full Mumford-Shah model (1.1) is in finding the smooth approximation u . This typically requires to solve some elliptic equations over each phase, e.g., $(\Delta - \mathbb{I})u = f$, where $\mathbb{I}$ is the identity operator and f is independent of u . A natural idea is to write $u = (\Delta - \mathbb{I})^{-1}f$, where $(\Delta - \mathbb{I})^{-1}$ is an integral operator, but does not have a closed form if the solution domain is not regular. This inspires us to search for some alternative integral smoothing operator and eliminate u in the Mumford-Shah model. One candidate at our disposal is the heat kernel:

$$\int K_{\sqrt{2t}}(\mathbf{x} - \mathbf{y})f(\mathbf{y})d\mathbf{y} = (K_{\sqrt{2t}} * f)(\mathbf{x}), \quad t > 0,$$

where t acts as a scale space variable, and

$$K_{\sigma}(\mathbf{x}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma^2}\right).$$

We incorporate this idea into the Mumford-Shah model and produce some preliminary results as below:

Figure 6.1: Preliminary results

The choice of the smoothing operator and its evaluation will be essential for this approach. This research topic is ongoing.

Bibliography

- [1] R.A. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.
- [2] M.K. Agoston. *Computer Graphics and Geometric Modelling: Implementation and Algorithms*. Springer, 2005.
- [3] S.M. Allen and J.W. Cahn. A microscope theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27(6):1085–1095, 1979.
- [4] L. Alvarez, P.L. Lions, and J.M. Morel. Image selective smoothing and edge detection by nonlinear diffusion II. *SIAM Journal on Numerical Analysis*, 29(3):845–866, 1992.
- [5] L. Ambrosio and V.M. Tortorelli. Approximation of functions depending on jumps by elliptic functions via Γ -convergence. *Communications on Pure and Applied Mathematics*, 13:999–1036, 1990.
- [6] L. Ambrosio and V.M. Tortorelli. On the approximation of functionals depending on jumps by quadratic, elliptic functions. *Bollettino della Unione Matematica Italiana*, 6-B, 105-123, 1992.
- [7] F. Andreu-Vallo, V. Caselles, and J.M. Mazon. *Parabolic Quasilinear Equations Minimizing Linear Growth Functionals*. Birkhauser, 2004.
- [8] C. Arzelá. Un’osservazione intorno alle serie di funzioni. *Accademia delle Scienze dell’Istituto di Bologna*, 142–159, 1882

-
- [9] G. Ascoli. Le curve limiti di una variet data di curve. *Atti della Reale Accademia Dei Lincei Memorie della Class di Scienze Fisiche Matematiche e Naturali*, 18(3):521–586, 1883.
 - [10] J.F. Aujol, G. Aubert, L.B. Feraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88, 2005.
 - [11] J.F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.
 - [12] N. Badshah and K. Chen. Multigrid method for the chan-veese model in variational segmentation. *Communications in Computational Physics*, 4(2):294–316, 2008.
 - [13] N. Badshah and K. Chen. Image selective segmentation under geometrical constraints using an active contour approach. *Communications in Computational Physics*, 7(4):759–778, 2010.
 - [14] E. Bae, J. Yuan, and X.C. Tai. Global minimization for continuous multiphase partitioning problems using a dual approach. *International Journal of Computer Vision*, 92(1):112–129, 2011.
 - [15] S. Basu, D.P. Mukherjee, and S.T. Acton. Implicit evolution of open ended curves. In *IEEE International Conference on Image Processing*, 1:261–264, 2007.
 - [16] B. Berkels, A. Rätz, M. Rumpf, and A. Voigt. Extracting grain boundaries and macroscopic deformations from images on atomic scale. *Journal of Scientific Computing*, 35(1):1–23, 2008.
 - [17] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.
 - [18] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
 - [19] A. Braides. *Approximation of Free-Discontinuity Problems*. Springer, 1998.

- [20] L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.
- [21] X. Bresson and T.F. Chan. Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Problems and Imaging*, 2(4):455–484, 2008.
- [22] X. Bresson, S. Esedoglu, P. Vanderghelynst, J.P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.
- [23] A. Brook, R. Kimmel, and N.A. Sochen. Variational restoration and edge detection for color images. *Journal of Mathematical Imaging and Vision*, 18(3):247–268, 2003.
- [24] E.S. Brown, T.F. Chan, and X. Bresson. Completely convex formulation of the Chan-Vese image segmentation model. *International Journal of Computer Vision*, 98(1):103–121, 2012.
- [25] E.S. Brown, T.F. Chan, and X. Bresson. A convex relaxation method for a class of vector-valued minimization problems with applications to Mumford-Shah image segmentation. *UCLA-CAM-report 10-43*, July, 2010.
- [26] M. Burger. A framework for the construction of level set methods for shape optimization and reconstruction. *Interfaces and Free Boundaries*, 5:301–329, 2003.
- [27] M. Burger, B. Hackl, and W. Ring. Incorporating topological derivatives into level set methods. *Journal of Computational Physics*, 194:344–362, 2004.
- [28] J.F. Cai, S. Osher, and Z.W. Shen. Split Bregman methods and frame based image restoration. *Multiscale Modeling and Simulation*, 8(2):337–369, 2009.
- [29] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
- [30] V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66:1–31, 1993.
- [31] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

-
- [32] F. Catté, P.L. Lions, J.M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29(1):182–193, 1992.
 - [33] H. Cenicerros and T.Y. Hou. An efficient dynamically adaptive mesh for potentially singular solutions. *Journal of Computational Physics*, 172:1–31, 2001.
 - [34] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.
 - [35] A. Chambolle. Total variation minimization and a class of binary MRF models. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 136–152. Springer, 2005.
 - [36] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.
 - [37] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
 - [38] T. F. Chan, K. Chen, and X. C. Tai. Nonlinear multilevel schemes for solving the total variation image minimization problem. In *Image processing based on partial differential equations*, 265–288. Springer, Berlin, 2007.
 - [39] T.F. Chan and S. Esedoglu. Aspects of total variation regularized L^1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.
 - [40] T.F. Chan, G.H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.
 - [41] T.F. Chan and J.H. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
 - [42] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

- [43] O. Christiansen and X. C. Tai. Fast implementation of piecewise constant level set methods. In *Image processing based on partial differential equations*, 289–308. Springer, Berlin, 2007.
- [44] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49:1–23, 1943.
- [45] D. Cremers. A multiphase levelset framework for variational motion segmentation. In *Scale Space and Variational Methods in Computer Vision*, 599–614, 2003.
- [46] G. Dal Maso. *Introduction to Γ -Convergence*. Birkhauser, 1993.
- [47] G. Aubert, P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Springer-Verlag, New York, 2006.
- [48] R. Deriche. Using Canny’s criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.
- [49] P. Destuynder, M. Jaoua, and H. Sellami. A dual algorithm for denoising and preserving edges in image processing. *Journal of Inverse and Ill-Posed Problems*, 15(2):149–165, 2007.
- [50] Y.Q. Dong and M. Hintermuller. Multi-scale total variation with automated regularization parameter selection for color image restoration. In *Scale Space and Variational Methods in Computer Vision*, 271–281. Springer, Berlin, 2009.
- [51] I. Ekeland and R. Tman. *Convex Analysis and Variational Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.
- [52] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, 1992.
- [53] K. Fan. Minimax theorems. *Proceedings of the National Academy of Sciences of the United States of America*, 39(1):42–47, 1953.
- [54] R.T. Farouki and C.A. Neff. Analytic properties of plane offset curves. *Computer Aided Geometric Design*, 7:83–99, 1990.

-
- [55] W.M. Feng, P. Yu, S.Y. Hu, Z.K. Liu, Q. Du, and L.Q. Chen. Spectral implementation of an adaptive moving mesh method for phase-field equations. *Journal of Computational Physics*, 220(1):498–510, 2006.
 - [56] A. Giacomini. Ambrosio-Tortorelli approximation of quasi-static evolution of brittle fractures. *Calculus of Variations and Partial Differential Equations*, 22(2):129–172, 2005.
 - [57] J. Gilles. Noisy image decomposition: A new structure, texture and noise model based on local adaptivity. *Journal of Mathematical Imaging and Vision*, 28:285–295, 2007.
 - [58] T. Goldstein, X. Bresson, and S. Osher. Geometric applications of the split Bregman method: Segmentation and surface reconstruction. *Journal of Scientific Computing*, 45:272–293, 2010.
 - [59] T. Goldstein and S. Osher. The split Bregman method for L^1 regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.
 - [60] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms: Fundamentals*. Springer, 1993.
 - [61] Y. Huang, M.K. Ng, and Y.W. Wen. A fast total variation minimization method for image restoration. *Multiscale Modeling and Simulation*, 7:774–795, 2008.
 - [62] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
 - [63] R. Kimmel, A. Amir, and A. Bruckstein. Finding shortest paths on surfaces using level set propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):635–640, 1995.
 - [64] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. In *Scale Space and Variational Methods in Computer Vision*, 150–162. Springer, Berlin, 2008.
 - [65] J. Lellmann and C. Schnörr. Continuous multiclass labeling approaches and algorithms. *SIAM Journal on Imaging Sciences*, 4(4):1049–1096, 2011.
 - [66] S. Leung and H. Zhao. A grid based particle method for evolution of open curves and surfaces. *Journal of Computational Physics*, 228(20):7706–7728, 2009.

-
- [67] C. Li, C. Xu, C. Gui, and M.D. Fox. Level set evolution without re-initialization: A new variational formulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CPVR)*, 1:430-436, 2005.
- [68] F. Li and M.K. Ng. Kernel density estimation based multiphase fuzzy region competition method for texture image segmentation. *Communications in Computational Physics*, 8(3):623–641, 2010.
- [69] F. Li, M.K. Ng, and C. Li. Variational fuzzy Mumford-Shah model for image segmentation. *SIAM Journal on Applied Mathematics*, 70(7):2750–2770, 2010.
- [70] F. Li, M.K. Ng, T.Y. Zeng, and L.C. Shen. A multiphase image segmentation method based on fuzzy region competition. *SIAM Journal on Imaging Sciences*, 3(3):277–299, 2010.
- [71] H.W. Li and X.C. Tai. Piecewise constant level set methods for multiphase motion. *International Journal of Numerical Analysis and Modeling*, 4:274–293, 2007.
- [72] J. Lie, M. Lysaker, and X.C. Tai. A binary level set model and some applications to mumford-shah image segmentation. *IEEE Transactions on Image Processing*, 15(5):1171–1181, 2006.
- [73] J. Lie, M. Lysaker, and X.C. Tai. A variant of the level set method and applications to image segmentation. *Mathematics of Computation*, 75(255):1155–1174, 2006.
- [74] C.X. Liu, F.F. Dong, S.F. Zhu, D.K. Kong, and K.F. Liu. New variational formulations for level set evolution without reinitialization with applications to image segmentation. *Journal of Mathematical Imaging and Vision*, 41(3):194-209, 2011.
- [75] J. Liu, X.C. Tai, H.Y. Huang, and Z.D. Huan. A fast segmentation method based on constraint optimization and its applications: Intensity inhomogeneity and texture segmentation. *Pattern Recognition*, 44(9):2093–2108, 2011.
- [76] B. Llanas and S. Lantarón. Edge detection by adaptive splitting. *Journal of Scientific Computing*, 46(3):486–518, 2011.

-
- [77] T. Lu, P. Neittaanmaki, and X.C. Tai. A parallel splitting up method and its application to Navier-Stokes equations. *Applied Mathematics Letters*, 4(2):25–29, 1991.
- [78] T. Lu, P. Neittaanmaki, and X.C. Tai. A parallel splitting up method for partial differential equations and its application to Navier-Stokes equations. *RAIRO Modélisation Mathématique et Analyse Numérique*, 26(6):673–708, 1992.
- [79] Z. Luo, L.Y. Tong, J.Z. Luo, P. Wei, and M.Y. Wang. Design of piezoelectric actuators using a multiphase level set method of piecewise constants. *Journal of Computational Physics*, 228:2643–2659, 2009.
- [80] W.Y. Ma and B.S. Manjunath. Edgeflow: a technique for boundary detection and image segmentation. *IEEE Transactions on Image Processing*, 9(8):1375–1388, 2000.
- [81] R. Malladi, J.A. Sethian, and B.C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.
- [82] R. March and M. Dozio. A variational method for the recovery of smooth boundaries. *Image and Vision Computing*, 15(9):705–712, 1997.
- [83] G.I. Marchuk. *Splitting and Alternating Direction Methods*. Volume I of Handbook of Numerical Analysis, Elsevier, 1990.
- [84] U. Massari and I. Tamanini. Regularity properties of optimal segmentations. *Journal für die reine und angewandte Mathematik*, 420:61–84, 1991.
- [85] E. Meinhardt, E. Zacur, A.F. Frangi, and V. Caselles. 3D edge detection by selection of level surface patches. *Journal of Mathematical Imaging and Vision*, 34(1):1–16, 2009.
- [86] B. Merriman, J.K. Bence, and S. Osher. Motion of multiple functions: a level set approach. *Journal of Computational Physics*, 112(2):334–363, 1994.
- [87] Y. Meyer. Oscillating Patterns in Image Processing and in Some Nonlinear Evolution Equations. In *The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, 2001.

- [88] C.A. Micchelli, L.X. Shen, and Y.S. Xu. Proximity algorithms for image models: denoising. *Inverse Problems*, available online since April 2011. DOI:10.1088/0266-5611/27/4/045009.
- [89] C.A. Micchelli, L.X. Shen, Y.S. Xu, and X.Y. Zeng. Proximity algorithms for the L1/TV image denoising model. *Advances in Computational Mathematics*, available online since October 2011. DOI: 10.1007/s10444-011-9243-y.
- [90] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus de l'Académie des Sciences*, 255:1897–2899, 1962.
- [91] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.
- [92] J.M. Morel and S. Salimini. Segmentation of images by variational methods: A constructive approach. *Reviews in Mathematical Universidad Complutense de Madrid*, 1:169–182, 1988.
- [93] J.M. Morel and S. Salimini. Segmentation d'images par méthode variationnelle: Une preuve constructive d'existence. *Comptes Rendus de l'Académie des Sciences*, 308(15):465–470, 1989.
- [94] J.M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Basel: Birkhäuser, 1994.
- [95] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.
- [96] M. Nikolova, T.F. Chan, and S. Esedoglu. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006.
- [97] S. Osher and R.P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.
- [98] S. Osher and J.A. Sethian. Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

-
- [99] N. Paragios, Y.M. Chen, and O. Faugeras. *Handbook of Mathematical Models in Computer Vision*. Springer, 2006.
 - [100] N. Paragios and R. Deriche. Geodesic active contours and level sets for detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):266–280, 2000.
 - [101] D. Peng, B. Merriman, S. Osher, H. Zhao, and M. Kang. A PDE-based fast local level set method. *Journal of Computational physics*, 155:410–438, 1999.
 - [102] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
 - [103] D.L. Pham and J.L. Prince. An adaptive fuzzy c-means algorithm for the image segmentation in the presence of intensity inhomogeneities. *Pattern Recognition Letters*, 20:57–68, 1998.
 - [104] 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)*, 810–817, 2009.
 - [105] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the Mumford-Shah functional. In *12th International Conference on Computer Vision (ICCV)*, 1133–1140, 2009.
 - [106] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*, 3(4):1122–1145, 2010.
 - [107] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *European Conference on Computer Vision (ECCV)*, 792–805, 2008.
 - [108] R.B. Potts. Some generalized order-disorder transformations. In *Proceedings of the Cambridge Philosophical Society*, 48:106–109, 1952.
 - [109] R.T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

- [110] Z.J. Rong, L.L. Wang, and X.C. Tai. Adaptive wavelet collocation methods for image segmentation using TV-Allen-Cahn type models. *Advances in Computational Mathematics*, available online since September 2011, DOI: 10.1007/s10444-011-9227-y.
- [111] H.L. Royden. *Real Analysis*. MacMillan Publishing Company, 1988.
- [112] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.
- [113] G. Sapiro and D.L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Transactions on Image Processing*, 5(11):1582–1586, 1996.
- [114] J.A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, 1999.
- [115] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In *Scale Space and Variational Methods in Computer Vision*, 464–476. Springer, Berlin, 2008.
- [116] J. Shah. A common framework for curve evolution, segmentation and anisotropic diffusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, 136–142, 1996.
- [117] J. Shen. A stochastic-variational model for soft Mumford-Shah segmentation. *International Journal of Biomedical Imaging*, 2006:2–16, 2006.
- [118] J. Shen, T. Tang, and L.L. Wang. *Spectral Methods: Algorithms, Analysis and Applications*. Springer, 2011.
- [119] J. Shen and X.F. Yang. An efficient moving mesh spectral method for the phase-field model of two-phase flows. *Journal of Computational Physics*, 228(8):2978–2992, 2009.
- [120] Y.Y. Shi, L.L. Wang, and X.C. Tai. Geometry of total variation regularized L^p -model. *Journal of Computational and Applied Mathematics*, 236(8):2223–2234, 2012.
- [121] P. Smereka. Spiral crystal growth. *Physica D: Nonlinear Phenomena*, 138(3-4):282–301, 2000.

-
- [122] S.M. Smith. Edge thinning used in the Susan edge detector. *Technical Report, TR95SMS5*, 1995.
- [123] G. Strang. On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, 5(3):506–517, 1968.
- [124] J.C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Society for Industrial and Applied Mathematics, 2004.
- [125] Y. Sun, P. Wu, G.W. Wei, and G. Wang. Evolution-operator-based single-step method for image processing. *International Journal of Biomedical Imaging*, 2006:1–28, 2006.
- [126] Y. Suzuki, T. Takayama, I.N. Motoike, and T. Asai. A reaction-diffusion model performing stripe-and spot-image restoration and its LSI implementation. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 90(1):20–29, 2007.
- [127] X.C. Tai and T.F. Chan. A survey on multiple level set methods with applications for identifying piecewise constant functions. *International Journal of Numerical Analysis and Modeling*, 1(1):25–47, 2004.
- [128] X.C. Tai, O. Christiansen, P. Lin, and I. Skjælaaen. Image segmentation using some piecewise constant level set methods with MBO type of projection. *International Journal of Computer Vision*, 73(1):61–76, 2007.
- [129] X.C. Tai and C. Wu. Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model. In *Scale Space and Variational Methods in Computer Vision*, 502–514. Springer, Berlin, 2009.
- [130] B. Tang, G. Sapiro, and V. Caselles. Color image enhancement via chromaticity diffusion. *IEEE Transactions on Image Processing*, 10(5):701–707, 2001.
- [131] W. Tao, F. Chang, L. Liu, H. Jin, and T. Wang. Interactively multiphase image segmentation based on variational formulation and graph cuts. *Pattern Recognition*, 43(10):3208–3218, 2010.
- [132] Z. Tari, J. Shah, and H. Pien. Extraction of shape skeletons from grayscale images. *Computer Vision and Image Understanding*, 66(2):133–146, 1997.

-
- [133] V. Toponogov. *Differential Geometry of Curves and Surfaces: A Concise Guide*. Birkhauser, 2006.
- [134] M. Upmanyu, R.W. Smith, and D.J. Srolovitz. Atomistic simulation of curvature driven grain boundary migration. *Interface Science*, 6:41–58, 1998.
- [135] L.A. Vese and T.F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.
- [136] L.L. Wang and Y. Gu. Efficient dual algorithms for image segmentation using TV-Allen-Cahn type models. *Communications in Computational Physics*, 9(4):859–877, 2011.
- [137] L.L. Wang, Y. Shi, and X.C. Tai. Robust edge detection using mumford-shah model and binary level set method. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, 291–301, Israel, 2011.
- [138] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.
- [139] G.W. Wei and Y.Q. Jia. Synchronization-based image edge detection. *Europhysics Letters*, 59(6):814–819, 2002.
- [140] J. Weickert, B. Romeny, and M.A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, 1998.
- [141] W. Wells, E. Grimson, R. Kikinis, and F. Jolesz. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 15:429–442, 1996.
- [142] C. Wu and X.C. Tai. Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM Journal on Imaging Sciences*, 3:300–339, 2010.
- [143] S.L. Xing and C.F. Shu. On the entropic regularization method for solving min-max problems with applications. *Mathematical Methods of Operations Research*, 46:119–130, 1997.

- [144] C. Zach, D. Gallup, J. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization Workshop (VMV)*, 243-252, 2008.
- [145] X.Q. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on Bregman iteration. *Journal of Scientific Computing*, 46(1):20–46, 2010.

List of Publications

- ① Li-Lian Wang and Ying Gu. Efficient dual algorithms for image segmentation using TV-Allen-Cahn type models. *Communications in Computational Physics*. Vol. 9, No. 4, 859-877, 2011.
- ② Ying Gu, Li-Lian Wang and Xue-Cheng Tai. A direct approach towards global minimization for multiphase labeling and segmentation problems. *IEEE Transactions on Image Processing*. Vol. 21, No. 5, 2399-2411, 2012.
- ③ Two other papers pertinent to this dissertation are in preparation.