Numerical Analysis of the Mumford-Shah and Mumford-Shah-Euler Functionals for Sphere-Valued Functions, and applications to Numerical Image Processing

Dissertation der Mathematischen Fakultät der
Eberhard-Karls-Universität Tübingen
zur Erlangung des Grades eines Doktors der Naturwissenschaften vorgelegt

by
Jonas Haehnle

2010

Tag der mündlichen Prüfung: 07. 07. 2010
Dekan: Prof. Dr. W. Knapp
1. Berichterstatter: Prof. Dr. A. Prohl
2. Berichterstatter: Prof. Dr. M. Rumpf

I am grateful to everyone who supported me throughout this dissertation and my education in general.

In particular, I would like to thank my adviser, Prof. Dr. Andreas Prohl of the University of Tübingen, for giving me the opportunity for this foray into the fields of numerical analysis and numerical image processing, for many helpful discussions, and for his constant encouragement.

I am also grateful to Prof. Dr. Andreas Prohl and Prof. Dr. Martin Rumpf (University of Bonn) for being the referees of this dissertation, and to Prof. Dr. Sören Bartels (University of Bonn) for his help with the coding.

Furthermore, I would like to thank all my colleagues form both numerical analysis groups at the University of Tübingen. I am particularly grateful to Erich Carelli, Dr. Ludwig Gauckler, Dr. Anton Prochel, and Dr. Markus Schmuck for discussions on mathematical and computational issues.

And last, not least, I would like to thank my friends and family for their unwavering support.

Abstract

The objective of this thesis is to find and analyse practical numerical algorithms for the minimisation and gradient-flows of the *Mumford-Shah* and *Mumford-Shah-Euler* functionals for unit vector fields.

The motivation for these questions is twofold: First, these are interesting model-problems combining non-convex functionals with a non-convex constraint, as an extension of existing works on harmonic maps to the sphere.

Second, bot functionals were originally introduced in image processing: The Mumford-Shah functional for *segmentation*, and the Mumford-Shah-Euler functional for *inpainting*; and the sphere-constraint can be used to implement the *chromaticity and brightness* colour model in this context.

In the first part of the thesis, two schemes for the minimisation of the Mumford-Shah functional for unit-vector fields are presented and discretised using first-order finite elements. The first scheme uses a projection approach to enforce the sphere-constraint. It works well in simulations, but we only have partial convergence results. The second scheme uses a penalisation approach, which only approximates the sphere-constraint, but allows for a complete proof of convergence.

In the second part of the thesis, two schemes for the gradient-flow of the Mumford-Shah-Euler functional for unit-vector fields are presented and discretised, again using first-order finite elements. The first scheme is an extension of the penalisation approach from part one, which again allows for a complete proof of convergence. The second scheme uses a Lagrange multiplier to enforce the sphere constraint, and we can again only present partial convergence results.

Both parts are concluded by simulations comparing the two corresponding algorithms with each other and presenting comparisons between the chromaticity and brightness and the conventional RGB colour model.

Zusammenfassung

Ziel dieser Arbeit ist es, praktikable numerische Minimierungs- und Gradientenfluss-Verfahren für das *Mumford-Shah* und das *Mumford-Shah-Euler* Funktional für Einheitsvektorfelder zu finden und zu analysieren.

Die Motivation für diese Problemstellung ist zweifach: Zum einen sind dies interessante Modellprobleme, die nichtkonvexe Funktionale mit einer nichtkonvexen Nebenbedingung kombinieren und als Erweiterung bestehender Arbeiten über Harmonische Abbildungen in die Sphäre gesehen werden können.

Zum anderen kommen beide Funktionale ursprünglich aus der Bildverarbeitung: Das Mumford-Shah Funktional wurde zur *Segmentierung*, das Mumford-Shah-Euler Funktional zum *Inpainting* vorgeschlagen; und die Sphärenbedingung kann in diesem Zusammenhang eine Anwendung im *Chromaticity and Brightness* Farbmodell finden.

Im ersten Teil der Arbeit werden zwei Minimierungsverfahren für das Mumford-Shah Funktional für Einheitsvektorfelder angegeben und mithilfe von finiten Elementen erster Ordnung ins Diskrete übertragen. Das erste Verfahren erhält die Sphärenbedingung mithilfe eines Projektionsansatzes. Es liefert überzeugende Simulationen, zur Konvergenz können aber nur Teilresultate präsentiert werden. Das zweite Verfahren verwendet einen Penalisierungsansatz, der die Sphärenbedingung nur approximiert, dafür aber eine vollständige Konvergenzanalyse zulässt.

Im zweiten Teil der Arbeit werden zwei Gradientenfluss-Verfahren für das Mumford-Shah-Euler Funktional für Einheitsvektorfelder vorgestellt und wieder mithilfe finiter Elemente erster Ordnung diskretisiert. Das erste Verfahren erweitert den aus dem ersten Teil bekannten Penalisierungsansatz, der auch hier eine vollständige Konvergenzanalyse erlaubt. Das zweite Verfahren verwendet einen Lagrange-Multiplikator, der eine exakte Einhaltung der Sphärenbedingung garantiert, für den aber wieder nur Teilresultate zur Konvergenz präsentiert werden können.

In Simulationen im Abschluss der beiden Teile werden jeweils die beiden vorgestellten Algorithmen untereinander vergleichen, sowie Bildverarbeitungs-Vergleiche zwischen dem Chromaticity and Brightness und dem herkömmlichen *RGB* Farbmodell angestellt.

Contents

Chapter 1

Introduction

In this chapter, we give a short introduction into the workings of the human eye, the basic concepts of colour models and image processing, as well as some background about the two functionals studied throughout the rest of this thesis, namely the Mumford-Shah and the Mumford-Shah-Euler functionals. General references are e.g. [60, 93, 34], in particular for Subsections 1.1, 1.2, and 1.3, respectively.

1.1 The Human Eye

The human eye acts similarly to a camera. The cornea and the lens act like a camera's lens to focus images. Between them, the iris/pupil controls the amount of light passing through, like the aperture in photography. Finally, the light is captured by the retina, which acts like a camera's film or image sensor.

The *cornea* is the the outer surface of the eye. It has no blood vessels and is therefore nurtured through marginal vessels and the fluids around it. The cornea's index of refraction is slightly higher than that of water, so that the change of index of refraction between the cornea and the air outside is the most significant of the whole human eye. Therefore many common defects of vision (myopia, hyperopia, or astigmatism) can be attributed to variations in the cornea's shape, and sometimes surgically corrected.

The *lens* is layered and flexible, with an index of refraction similar to the cornea's. Its shape is controlled through the ciliary muscles, and it allows the eye to accommodate; i.e., focus. The lens' flexibility decreases with age so that typically, after an age of about 50, it no longer allows to focus on near objects (presbyopia). The lens' optical density also increases with age, increasing the amount to which short-wavelength (blue and violet) energy is absorbed and scattered (chromatic adaptation generally makes people unaware of this, but it is apparent when doing colour matching tests).

The volume between the cornea an the lens is occupied by the *aqueous humour*, whose viscosity and index of refraction are similar to water, while the space between the lens an retina is occupied by the *vitreous humour* with similar index of refraction, but higher viscosity (more like gelatin). Both fluids have somewhat higher than air pressure to keep the flexible eyeball in shape.

The *iris* is the pigmented muscle controlling the pupil size. Eye colour is determined by the concentration and distribution of melanin in the iris. The pupil size determines the

amount of light that passes inside. The pupil's diameter varies from approximately 3mm to 7mm, which corresponds to an approximately fife-fold change in retina illumination (which is not sufficient to explain human vision over illumination levels varying over 10 orders of magnitude).

The *retina* is a thin layer of cells at the back of the eye, which contains the photosensitive cells responsible for gathering the image projected onto the retina by the optical elements of the eye. These neurons, called *rods* and *cones* according to their typical shape, transform the optical information into chemical and electrical signals that are then processed by a network of cells (*horizontal*, *bipolar*, and *amacrine* cells), that sit on top of the photoreceptors (towards the light, which is why they need to be transparent) and led to the brain by the optic nerve. This network already acts as an image computer, compressing, among other things, the information from 130 million photoreceptors to signals in about one million ganglion cells. Rods are effective at much lower luminance levels than cones. There is only one type of rod receptor with a peak sensitivity for light of about 510 nm wavelength. On the other hand, there are three types of cones: L, M, and S cones most sensitive to long ($\sim$ 570 nm), medium ($\sim$ 540 nm), and short ($\sim$ 430 nm) wave lengths. These cones are also referred to as red, green, and blue, which is a bit misleading, since their response curves are strongly overlapping. The relative populations of cones are approximately 12:6:1 (L:M:S). In total, there are about 120 million rods and about 7 million cones per retina. Usually, single cones feed into ganglion cell signals, while cones are pooled over hundreds of receptors to increase sensitivity.

According to the modern theory of colour vision, the first stage of colour vision (the receptors) is trichromatic (LMS). After this stage, the neurons of the retina encode these three images into opponent signals: luminance ($L+M+S$), red-green ($L-M+S$), and yellow-blue ($L+M-S$) signals are calculated, which decorrelate the colour information, allowing more efficient transmission and reducing difficulties with noise. This opponent colour nature of signal transmission is the reason why we never perceive a colour as red-green or yellow-blue.

Visual experiments with stimuli that vary sinusoidally across space show that the spatial contrast sensitivity of human vision is very different in luminance and chromaticity: Luminance contrast sensitivity is zero at zero cycles per degree (the visual system tends to be insensitive to uniform fields), has its peak at about five cycles per degree and approaches zero again at about 60 cycles per degree (no detail can be resolved any more). Blue-yellow and red-green contrast sensitivity, on the other hand, is low-pass in nature; i.e., low frequencies can be detected, but contrast sensitivity approaches zero already at or even before five cycles per degree. This is why substantial colour subsampling can be effectively used to compress images, without much visible effect, see also below.

Behind the retina is the *pigmented epithelium*, a dark layer that absorbs light that passes through the retina and prevents reflections.

The *fovea* is the area of highest spatial and colour resolution on the retina. This is achieved by a very high density of cones, while rods are completely absent (which is why dim stars can be more easily detected when not looking right at them: the rods outside the fovea are more light sensitive). It covers a visual angle of about two degrees in the central field of view. It is protected from intense short-wavelength energy by a yellow

filter called *macula*, the density of whose pigment strongly varies from person to person (and sometimes eye to eye), representing a major source of variability in colour vision.

The *optic nerve* consists of approximately 1.2 million *retinal ganglion cell axons*. The area taken up by the optic nerve is not photo receptive and is called the *blind spot*.

1.2 Colour Models

Most real-world images are colour images, and even in greyscale images, e.g. from X-rays or space probes, false colours are often used to encode additional information or to make differences in shades of grey (of which humans can only detect about 20 to 30) more visible through colours (of which we can distinguish hundreds).

In order to process colour images, one needs a mathematical representation of colours, a so called *colour model*. This could e.g. be a description of how many units of any of the three primary colours a painter needs to mix for a given colour, resulting in a colour vector representing the values for red, yellow, and blue. If associated with a description of how exactly to interpret the components of colour vectors, a colour model leads to a *colour space*.

In digital cameras and computer screens, *RGB* is the most common colour model. Here, brightness values of red, green, and blue signals are stored individually. This is to some extent similar to the way the human eye detects colours through separate cones for long, medium, and short wavelengths (LMS). However, there are significant differences: The sensitivity of the different cones is strongly overlapping, while RGB channels are not, and the eye additionally employs rods (which are much more sensitive in low light) for brightness detection. In fact, RGB was not devised to work like the human eye, but to be well-adapted to the way camera sensors and display phosphors work. RGB is an additive colour model, meaning that mixing the RGB channels is like combining light sources of different colours. So, combining red and green produces yellow, and combining all three colours in equal amounts gives various shades of grey, zero red, green, and blue giving black, full red, green, and blue giving white. RGB can be displayed as a cube inside the three-dimensional Cartesian space, each axis representing one of the primary colours red, green, and blue, while the diagonal between the origin (no red, green, and blue) and full red, green, and blue represents all shades of grey. The most common implementation of RGB is 24-bit, i.e. 8 bits or 256 discrete levels of colour are stored per channel: $R, G, B \in \{0, \ldots, 255\}$. Some professional uses require more (typically 16) bits per channel, resulting in the same colour range but much higher density of colours (finer shades of colours), and higher memory use.

CMY is the complementary colour model to RGB. This abbreviation stands for cyan, magenta, and yellow, and this model is primarily used in print. It is a subtractive colour model, which means that it works like paint: Adding cyan and yellow gives green. Adding cyan, magenta, and yellow in equal amounts should give grey, if small amounts are taken, and black for full amounts. In practice, however, the result usually turns out to be some shade of brown. To make up for this, a black component is usually added, so CMY becomes *CMYK* (cyan, magenta, yellow, and black). Another reason for the addition of

black is that text is usually black, so black print is used a lot, whence using one unit of black ink instead of three units of coloured ink makes sense.

The RGB and CMYK colour models are both hardware-oriented, and corresponding colour spaces are device-dependent, and used in very different devices. Therefore there is no simple, general, or exact conversion between the two. Colour management systems employing device-dependent colour profiles are usually used for this conversion. Also, the RGB and CMYK colour models are better adapted to screens and printers, respectively, than to the human perception of colour.

One attempt at solving these issues are the *HSV* (hue, saturation, value), *HSI* (hue, saturation, intensity), and *HSL* (hue, saturation, lightness, also called *HLS*) colour models, which are very similar to each other. They all try to model the artist's concepts of tint, shade, and tone. Hue is the colour as defined by wavelength (red or yellow); saturation is the amount of that colour that is present (red or pink); the third axis (lightness, intensity, or value) represents the amount of light present (dark red or light red). These three spaces can be plotted as a double cone, in which the central axis represents the progression from black through all shades of grey to white. The distance from this axis represents the saturation, while the direction is the hue.

These models have many advantages for processing and understanding images, because colour information is separated into components that to some extent correspond to the human visual system's perception, and because chromatic information and brightness/luminance can be treated separately, which allows algorithms to give different weight to them. This makes sense because the human visual system takes most geometric cues from an images luminance, whence colour subsampling has a long tradition, e.g. in television standards or still image compression. A further reason to separate chromatic and luminance information is the fact that chromatic noise is usually much more prominent and disturbing in images from digital cameras, which has to do with the sensor technology commonly employed (analogue film does not suffer from this effect): Photocells only count photons, they cannot detect colour. Therefore filters need to be used — one filter per *pixels* (picture element). A square of four pixels on a digital sensor usually has one pixel sensitive to red, one sensitive to blue, and two sensitive to green (*Bayer matrix*). The remaining information needs to be interpolated.

Mathematically, however, the HSV / HSI / HSL colour models are rather awkward (see also the conversions below): Hue cycles through the angles from 0° to 360°, and then wraps around; and because of the double cone shape, increasing intensity or luminance can alter the saturation. A conversion from RGB can be done by defining

$$\begin{aligned} M &= \max\{R, G, B\}, \\ m &= \min\{R, G, B\}, \\ C &= M - m. \end{aligned}$$

Then

$$H = \begin{cases} \text{undefined} & \text{if } C = 0, \\ 60^\circ \left(\dfrac{G-B}{C} \mod 6 \right) & \text{if } M = R, \\ 60^\circ \left(\dfrac{B-R}{C} + 2 \right) & \text{if } M = G, \\ 60^\circ \left(\dfrac{R-G}{C} + 4 \right) & \text{if } M = B, \end{cases}$$

$$\begin{aligned} I &= \frac{R+G+B}{3}, \\ V &= M, \\ L &= \frac{M+m}{2}, \end{aligned}$$

$$\begin{aligned} S_{HSV} &= \begin{cases} 0 & \text{if } C = 0, \\ \dfrac{C}{V} = \dfrac{M-m}{M} & \text{otherwise}, \end{cases} \\ S_{HSI} &= \begin{cases} 0 & \text{if } C = 0, \\ 1 - \dfrac{m}{I} & \text{otherwise}, \end{cases} \\ S_{HSL} &= \begin{cases} 0 & \text{if } C = 0, \\ \dfrac{C}{2L} = \dfrac{M-m}{M+m} & \text{if } L \leq \dfrac{1}{2}, \\ \dfrac{C}{2-2L} = \dfrac{M-m}{2-(M+m)} & \text{if } L > \dfrac{1}{2}. \end{cases} \end{aligned}$$

A geometrically preferable colour model with similar properties to HSV, which also allows chromatic information and brightness to be treated separately, is the spherical *CIELAB*, or $L^*a^*b^*$ model, defined in 1976 by the *Commission Internationale d'Éclairage* (CIE). L^* is the greyscale axis (luminance), and is usually identified with the vertical axis in the unit ball. a^* and b^* are orthogonal axes defining colour and saturation. Here, $+a^*$ is red, $-a^*$ is green, $+b^*$ is yellow, and $-b^*$ is blue. This offers a compromise between the simpler RGB and the more physiological HSI. CIELAB is considered to be perceptually uniform; i.e., just-detectable visual differences should have a constant distance within the space. It is derived from the *CIE XYZ* colour space from 1931, and can be obtained as follows:

$$\begin{aligned} X &= 0.412453R + 0.357580G + 0.180423B \\ Y &= 0.212671R + 0.715160G + 0.072169B \\ Z &= 0.019334R + 0.119193G + 0.950227B. \end{aligned}$$

Then the L*a*b* values are

$$\begin{aligned} L^* &= 116 f\left(\frac{Y}{Y_n}\right) - 16 \\ a^* &= 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \\ b^* &= 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right), \end{aligned}$$

where f is defined as

$$f(q) := \begin{cases} q^{\frac{1}{3}} & \text{if } q > \left(\frac{6}{29}\right)^3, \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 q + \frac{4}{29} & \text{otherwise,} \end{cases}$$

and X_n, Y_n, Z_n are calculated for a reference white point depending on the illumination of the scene.

Finally, the *CB* colour model (*Chromaticity and Brightness*) separates colour into a spherical chromaticity $\mathbf{c}$ (living on part of the surface of the unit ball) and a separate scalar brightness component b. It is geometrically pleasing and easily obtained from RGB:

$$\begin{aligned} b &= \sqrt{R^2 + G^2 + B^2}, \\ \mathbf{c} &= \frac{(R, G, B)}{b} \in \mathbb{S}^2 \subset \mathbb{R}^3. \end{aligned}$$

(Note that here the chromaticity component does not cover the whole sphere, but merely a spherical triangle spanned by the points where the Cartesian axes intersect with the unit sphere centred in the origin.) While this chromaticity component is again less intuitive than hue and saturation, it provides a coupling between these two which seems to be useful in image processing: In [31] it is demonstrated, that this model is advantageous to RGB and to HSV in colour image denoising (using total variation, see below for an explanation of these terms). The reason seems the more natural coupling between dimensions in the chromaticity component, as the authors achieve results similar to CB with HSV by introducing a coupling between hue and saturation, while straightforward channel-wise RGB application lags far behind.

To get a more "natural" representation of colours in the CB colour model, it was proposed in [71] to apply a linear transformation to the XYZ-values obtained from RGB, before doing the above conversion to CB:

$$\begin{aligned} \tilde{R} &= 0.922X - 0.853Y - 0.069Z \\ \tilde{G} &= 0.173X + 0.189Y - 0.036Z \\ \tilde{B} &= 0.344X + 0.346Y + 0.123Z. \end{aligned}$$

(This also increases the area of the sphere covered by the chromaticity component.) This is the model we are going to refer to as CB.

1.3 Mathematical Image Processing

Image processing developed as a sub-field of electrical engineering, as a form of signal processing, applying standard signal-processing techniques to an input that happened to be an image. The term image processing is nowadays usually used synonymously with digital image processing, which means the use of computer algorithms to process digital images. This kind of image processing was developed in the 1960s at a few specialised laboratories primarily in the United States (Bell Labs, MIT and others). Specialised and very expensive hardware was used, for applications like satellite imagery and medical imaging. With the advent of more powerful and cheaper personal computers, the use of digital image processing proliferated, and has now indeed become the most common form of image processing. Most of the more involved image processing tasks, like classification, feature extraction, or pattern recognition, would not be possible by analogue means, but the more elementary tasks also profit from the use of more complex algorithms, that digital image processing allows.

Greyscale images can be thought of as maps $g : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}^2$ is the image domain, and $0 \leq g(\mathbf{x}) \leq 1$ for $\mathbf{x} \in \Omega$, 0 and 1 representing e.g. black and white, respectively. Colour images, e.g. in the RGB model, can be represented as vector-valued maps $\mathbf{g} : \Omega \to \mathbb{R}^3$, with $\mathbf{g}(\mathbf{x}) \in [0,1]^3 \subseteq \mathbb{R}^3$ for $\mathbf{x} \in \Omega$, the three axes representing red, green, and blue.

Mappings and vector spaces are the way we understand colours mathematically, but since computers and other electronic devices have limited resources, we need to discretise images in order to be able to perform actual computations on them. So the image domain is represented by a discrete set of points (pixels), each of which is associated a colour or greyscale value. These values, again, need to be discrete, e.g. integer numbers between 0 and 255, and are stored as matrix, or *raster map*.

Historically, greyscale images were the first to be electronically processed. And with the advent of colour images, the most natural thing to do was to apply the well-established algorithms for greyscale images to each colour-component (e.g. R,G, and B) separately. But as stated above, some coupling between colours seems to be advantageous, as is separation of brightness and colour, among other reasons because the image brightness usually contains most of the geometrical information, while e.g. the most disturbing kind of noise in digital images usually is chromatic in nature, and because the human visual system apparently does something similar.

Let us mention some prominent tasks in digital image processing.

- *Denoising* tries to improve noisy images by smoothing regions which do not contain important features, thereby removing insignificant information, sensor noise, etc., while hopefully preserving important features like edges.

- *Compression*, like denoising, simplifies images by neglecting less important features. Here, however, the goal is to lose as little visual information as possible, while saving as much storage space as possible.

- *Segmentation* tries to partition images into smooth regions, so that discontinuities only remain between segments. This task is almost trivial to the human visual

system, but remains one of the big problems in image processing. It is also a task that bridges low-level image processing and higher-level tasks such as scene recognition, image classification, image understanding, and artificial intelligence.

- *Inpainting* fills holes in images degraded e.g. by lossy transmission or age, or removes unwanted objects from images (like tourists in front of monuments or functionaries who have fallen from favour in official photographs). The word inpainting is an artistic synonym for image interpolation and stems from traditional museum image restoration. In fact, digital image inpainting has been used successfully to help restore historic frescoes.

- *Matching* or *registration* assigns complementary information contained in different images to one geometric reference model, e.g. in medical image processing, since clinical diagnosis and therapy are often supported by several types of images of the same patient.

It can sometimes make sense to perform several of these tasks at once, since e.g. segmenting a noisy image is difficult without denoising, while denoising first and segmenting afterwards might result in a loss of valuable information that could be avoided by doing both tasks simultaneously.

Some of the techniques of digital image processing are mentioned below (described for greyscale images, for simplicity).

- *Fourier Analysis* is one of the favoured tools of classical signal processing, and has been successfully applied to many image processing tasks; e.g. to compression algorithms. An image $g : (0,1)^2 \to \mathbb{R}$ (with periodic extension) can be encoded into its Fourier coefficients

$$c_{\mathbf{n}} = \int_\Omega g(\mathbf{x}) \exp(2\pi i \mathbf{x} \cdot \mathbf{n}) \mathrm{d}\mathbf{x}, \qquad \mathbf{n} \in \mathbb{Z}^2,$$

  which, in the case of a discretised image, becomes a discrete Fourier transform, efficiently implementable through *fast Fourier transform (FFT)*. However, the harmonic basic functions used by the Fourier transform are not ideal for discontinuous features like edges. Dirac's delta function is an extreme example: The coefficients $|c_{\mathbf{n}}|$ never decrease, so that this simple image generates a huge amount of data in the frequency domain.

- *Wavelet Analysis* (or more recently *Ridgelet Analysis*) solves this problem, basically by using more localised (fast decaying) basis functions that can be purpose-built for special applications.

- In *Stochastic Modelling*, in particular *Bayesian modelling*, a conditional a posteriori probability $p(F|Q)$ is maximised, Q denoting the observed image, F the "ideal image" generating Q, and p a probability distribution describing the specific model used.

- *Variational Methods* minimise an a priori "energy" E, which in this case describes the specific model in use, and which is related to the probability distribution above. The Mumford-Shah and Mumford-Shah-Euler Functionals discussed below are examples of such energies.
- *Partial Differential Equations (PDEs)* allow techniques developed in classical mathematical physics to be applied to image processing. Through calculus of variations, many variational problems can be formulated as PDEs, so that these approaches are again closely related.

1.4 The Mumford-Shah Functional

As mentioned above, image processing through variational methods minimises an a priori energy E encoding the model, so that the processed image u is given by

$$u = \operatorname*{argmin}_v E(v).$$

E is supposed to assign small values to "good" images, and large values to "bad" images.

The Mumford-Shah Functional is one such energy, which was proposed in 1989 by David Mumford and Jayant Shah in their paper [86]. For a given greyscale image $g : \Omega \to \mathbb{R}$, it addresses the problem of denoising and segmentation in a very natural way:

Maybe the simplest denoising energy is

$$E(u) := \int_\Omega |\nabla u|^2 \mathrm{d}\mathbf{x}.$$

This energy assigns large values to images with large gradients. Minimising it therefore minimises gradients, and makes the image smooth, indeed, too smooth: All features are smoothed indiscriminately. What we would like is to exempt an *edge set* K from the smoothing, which then is also an unknown:

$$E(u, K) := \int_{\Omega \setminus K} |\nabla u|^2 \mathrm{d}\mathbf{x}.$$

Of course, this energy does not need to have anything to do with the given image g. We therefore add a *fidelity term*:

$$E(u, K) := \int_{\Omega \setminus K} |\nabla u|^2 \mathrm{d}\mathbf{x} + \int_\Omega (u - g)^2 \mathrm{d}\mathbf{x}.$$

Yet, the functional above still has a trivial minimiser: $u = g$, and $K = \Omega$, meaning the whole image is full of edges. To avoid this problem, we need to penalise edges, or edge length (since we think of edges as lines). We do this by adding the one-dimensional Hausdorff measure of K to the energy:

$$E(u, K) := \int_{\Omega \setminus K} |\nabla u|^2 \mathrm{d}\mathbf{x} + \int_\Omega (u - g)^2 \mathrm{d}\mathbf{x} + \mathcal{H}^1(K)\,.$$

Finally, we add coefficients which will later allow us to balance the effects of these three terms:

$$E(u,K) := \frac{\gamma}{2}\int_{\Omega\setminus K}|\nabla u|^2\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_{\Omega}(u-g)^2\mathrm{d}\mathbf{x} + \alpha\mathcal{H}^1(K)\,,$$

which is to be minimised for all closed sets $K \subset \Omega$, and functions $u \in H^1(\Omega \setminus K)$; together with $g \in L^2(\Omega)$ these are the minimum requirements that give meaning to the above expression.

For this kind of problem, involving a functional depending on a variable set of discontinuities (edges, in our case), the term *free discontinuity problems* was introduced by De Giorgi in [43].

We remark that non-trivial images cannot be global Sobolev functions: Gradients of non-trivial images cannot be square integrable over the whole domain, since edges are jumps in image intensity, whence gradients of images with edges must be allowed to be unbounded on a closed edge set K (therefore $u \in H^1(\Omega \setminus K)$ above). This concept can be formalised in the framework of *functions of bounded variation*, which are functions whose distributional derivative is a measure with bounded total variation, see Section 2.2.1 for details.

Some results concerning existence of solutions and approximations to this functional are collected in Section 2.1.

1.5 The Mumford-Shah-Euler Functional

In image inpainting, large chunks of data inside the inpainting domain are missing or unreliable. For reasons of compatibility with other publications, we now call the edge set Γ, the image domain R, and the inpainting domain $K \subset R$. We want to fill this part of the image with data based on the information around it and based on our a priori model of what a "good" image should look like. For this, we may not need the information of the whole image, but only of a band $B \subset R \setminus K$ around the inpainting domain. The domain of computation could then be restricted to these two: $\Omega := K \cup B$.

Obviously, since there is no reliable information inside the inpainting domain, the fidelity term should only be regarded outside it; i.e., only inside B.

So if we apply the Mumford-Shah functional to this problem, it looks like this:

$$E(u,\Gamma) := \frac{\gamma}{2}\int_{\Omega\setminus\Gamma}|\nabla u|^2\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_{B}(u-g)^2\mathrm{d}\mathbf{x} + \alpha\mathcal{H}^1(\Gamma)\,.$$

Since this energy always favours the shortest possible edges (everything else being equal), a bar with a missing part in the middle, that is longer than the width of the bar, will be completed as two bars with empty space between them, even though this does not seem to be the most natural completion, see Figure 1.1.

Another example is a round cake with a piece missing: It will be completed with a straight line instead of a round one, see Figure 1.2 (and Example 3.1 in Section 3.6).

Introducing an *edge curvature* term in addition to the existing edge length term should solve this problem (the kinks in the boundaries of the two separate bars have very high curvature and should therefore make this unnatural inpainting very "expensive").

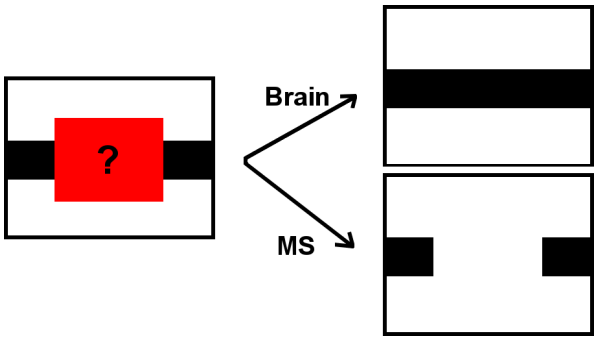

Figure 1.1: Image of an incomplete bar, with a natural and not so natural inpainting.

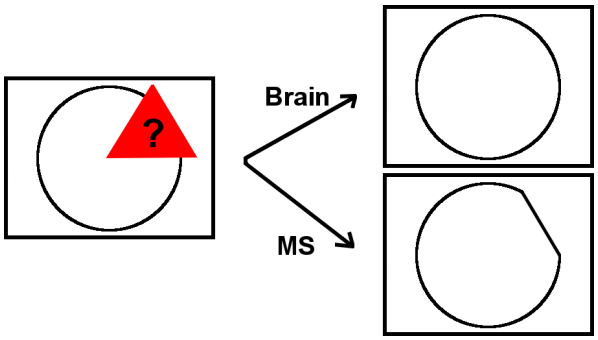

Figure 1.2: Image of an incomplete cake, with a natural and not so natural inpainting.

The *Mumford-Shah-Euler* model, which was proposed in [57], is based on exactly this idea (the name stems from Euler's elastica curve model):

$$E(u,\Gamma) := \frac{\gamma}{2}\int_{\Omega\setminus\Gamma} |\nabla u|^2 \mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_B (u-g)^2 \mathrm{d}\mathbf{x} + \int_\Gamma \left(\alpha + \beta|H|^2\right) \mathrm{d}\mathcal{H}^1(\mathbf{x})\,,$$

with H denoting the (mean) curvature of Γ. This functional is to be minimised for all $\Gamma \subset \Omega$ closed and C^2, and $u \in H^1(\Omega \setminus \Gamma)$. The new term (Euler term) with coefficient β penalises the curvature of Γ, and it corresponds to the *Willmore energy*.

Adding an edge curvature term represents a fundamental change in the functional. Edge length and curvature inside the inpainting domain now need to be balanced against each other. Therefore, depending on the weights in front of these terms, there can now be stable edge circles even in the absence of large gradients of the image u (as local minima of the energy): decreasing the radius r would decrease edge length but increase edge curvature (which behaves like $1/r$), compare Section 3.6. With the Mumford-Shah functional, less edges is always better, as long as the image u does not call for edges through a large gradient.

1.6 The Sphere Constraint

The two functionals introduced above, the Mumford-Shah, and the Mumford-Shah-Euler functional, are both non-convex, which introduces major difficulties when trying to minimise them.

In order to process colour images in the chromaticity and brightness colour model, we in addition have to introduce a non-convex constraint, namely the sphere-constraint (for the chromaticity), in which we are also interested from a purely theoretical point of view: The Mumford-Shah and Mumford-Shah-Euler functionals for sphere-valued functions are an exciting extension of the problem of sphere-valued harmonic maps, which has been intensely studied in the past.

For $\Omega \subseteq \mathbb{R}^d$, a map $\mathbf{u} \in H^1(\Omega, \mathbb{R}^m)$ is called *(weakly) harmonic*, if it is stationary for Dirichlet's energy

$$\frac{1}{2} \int_\Omega |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x},$$

or equivalently, if it satisfies the Euler-Lagrange equation

$$-\Delta \mathbf{u} = 0$$

(in the distributional sense). See [98] for a survey on the evolution of harmonic maps in the more general setting of Riemannian manifolds. Other sources for the theory of harmonic maps are in particular [94, 95], as well as [69, 58, 18, 77, 76, 95, 97, 68], and the book [99].

If $\mathbf{u} \in H^1(\Omega, \mathbb{R}^m)$ with $|\mathbf{u}| = 1$ almost everywhere is a stationary point of Dirichlet's energy, then it is a harmonic map to the sphere. Such maps are a nice prototype problem for harmonic maps to more general surfaces or manifolds, as well as being of interest to applications in image processing, liquid crystal theory, and micromagnetism, which explains why this special case has received much attention both analytically and numerically (e.g. [95, 58, 77] and [37, 38, 78, 1, 10, 12, 11, 6], respectively).

For iterative algorithms approximating harmonic maps to the sphere, the sphere constraint is a major difficulty. Three simple strategies for solving it immediately come to mind:

(1) *Projection*: Why not just solve the unconstrained problem an then project the solution to the sphere? This simple idea often works in practice, but it can be shown that in some cases Dirichlet's energy may grow because of the projection. However, variations of this idea have been successfully used in [37, 38, 78, 1, 10]. In particular, [1] uses a nice modification of the projection idea, that ensures energy decrease: Find an update $\mathbf{w}$ such that $\mathbf{u} - \mathbf{w}$ is stationary for Dirichlet's energy under the much simpler, linear constraint $\mathbf{u} \perp \mathbf{w}$ (i.e., $\mathbf{u} \cdot \mathbf{w} = 0$), and then project $\mathbf{u} - \mathbf{w}$ to the sphere. We shall apply this to the Mumford-Shah functional in Chapter 2. This idea is specific to the sphere as a target.

(2) *Penalisation*: If we add a term of the form $\frac{1}{\varepsilon} \int_\Omega \left(1 - |\mathbf{u}|^2\right)^2 \mathrm{d}\mathbf{x}$ to Dirichlet's energy, then the $|\mathbf{u}|$ will approximate 1 ever more closely for $\varepsilon \to 0$. While this no longer ensures an exact sphere constraint for actual calculations (where always $\varepsilon > 0$), the

penalisation idea has the advantage of working for more general targets. Related strategies have been considered in [52, 9, 8, 91, 80, 79], and we shall apply this idea to both the Mumford-Shah and the Mumford-Shah-Euler functional in Chapters 2 and 3, respectively.

(3) *Lagrange Multiplier*: Instead of basing our scheme directly on Dirichlet's energy, why not use the Euler-Lagrange equation? Of course, we also have to incorporate the sphere into the equation, which, for a harmonic map $\mathbf{u} : M \to N$ between Riemannian manifolds (N compact), can be expressed as $\Delta_M \mathbf{u} \perp T_{\mathbf{u}} N$, meaning that the Laplace-Beltrami operator of $\mathbf{u}$ on M is orthogonal to the tangent space of N at $\mathbf{u}$. In our case, $M = \mathbb{R}^d$, and $N = \mathbb{S}^{m-1} = \{\mathbf{x} \in \mathbb{R}^m : |\mathbf{x}| = 1\}$, and for something to be orthogonal to $T_{\mathbf{u}}\mathbb{S}^{m-1}$, it must be parallel to $\mathbf{u}$ itself; i.e.,

$$-\Delta \mathbf{u} = \lambda \mathbf{u}, \quad |\mathbf{u}| = 1$$

where λ is the Lagrange multiplier for the sphere constraint, and can be worked out to be $\lambda = |\nabla \mathbf{u}|^2$. This strategy has been used in [11, 6] for harmonic maps to spheres of both constant and varying radii (it also works for more general targets), and we are going to apply it to the Mumford-Shah-Euler functional in Chapter 3.

A version of this strategy that is specific to the sphere as a target was explored in [12]: By the properties of the cross product, the Euler-Lagrange equation given above is formally equivalent to

$$\mathbf{u} \times (\mathbf{u} \times \Delta \mathbf{u}) = 0,$$

which can also be used as a starting point for iterative algorithms.

1.7 Approximations of both Functionals

To represent the edge set K in the Mumford-Shah functional, we use an elliptic *phase field* functional suggested by Ambrosio and Tortorelli in [4, 5]. Let us repeat the Mumford-Shah functional:

$$E(u, K) = \frac{\gamma}{2} \int_{\Omega \setminus K} |\nabla u|^2 \mathrm{d}\mathbf{x} + \frac{\lambda}{2} \int_{\Omega} (u - g)^2 \mathrm{d}\mathbf{x} + \alpha \mathcal{H}^1(K)\,.$$

The idea is to introduce a *phase function* s such that $\{s = 0\} \approx K$. This is achieved by modifying the first term like this:

$$\frac{\gamma}{2} \int_{\Omega} s^2 |\nabla u|^2 \mathrm{d}\mathbf{x},$$

so $|s|$ is forced to be small whenever $|\nabla u|$ is large. As suggested in [84], the last term can be approximated in the sense of Γ-convergence (see Section 2.2.1 for details) by

$$\alpha \int_{\Omega} \left(\varepsilon \, |\nabla s|^2 + \frac{1}{4\varepsilon} (1 - s)^2 \right) \mathrm{d}\mathbf{x}.$$

Basically, the first part of this integral penalises phase transitions of s, while the second part forces s to be 1 wherever possible (for $\varepsilon \to 0$). Finally, the first term can be modified to lead to an elliptic functional. This then leads to the Ambrosio-Tortorelli functional:

$$\begin{aligned} E_\varepsilon(u,s) := \frac{\gamma}{2}\int_\Omega \left(s^2 + k_\varepsilon\right)|\nabla u|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega (u-g)^2 \mathrm{d}\mathbf{x} \\ + \alpha \int_\Omega \left(\varepsilon\,|\nabla s|^2 + \frac{1}{4\varepsilon}(1-s)^2\right)\mathrm{d}\mathbf{x} \end{aligned}$$

for $u, g \in H^1(\Omega)$, $s \in H^1(\Omega,[0,1])$, $0 < \varepsilon, k_\varepsilon \ll 1$, and $k_\varepsilon = o(\varepsilon)$. Ambrosio and Tortorelli showed Γ-convergence of $E_\varepsilon(u,s)$ to a weak formulation of $E(u,K)$ for $\varepsilon \to 0$.

Now, let us repeat the Mumford-Shah-Euler functional:

$$E(u,\Gamma) := \frac{\gamma}{2}\int_{\Omega\setminus\Gamma} |\nabla u|^2 \mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_B (u-g)^2 \mathrm{d}\mathbf{x} + \int_\Gamma \left(\alpha + \beta|H|^2\right)\mathrm{d}\mathcal{H}^1(\mathbf{x})\,.$$

As conjectured in [44], and proved in [92], the last term can again be approximated in the sense of Γ-convergence (up to a constant), leading to the functional

$$\begin{aligned} E_\varepsilon(u,s) :=& \frac{\gamma}{2}\int_\Omega \left(s^2 + k_\varepsilon\right)|\nabla u|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_B (u-u_0)^2 \,\mathrm{d}\mathbf{x} \\ &+ \alpha \int_\Omega \left(\frac{\varepsilon}{2}|\nabla s|^2 + \frac{1}{\varepsilon}\psi(s)\right)\mathrm{d}\mathbf{x} + \frac{\beta}{\varepsilon}\int_\Omega \left(\frac{1}{\varepsilon}\psi'(s) - \varepsilon\Delta s\right)^2 \mathrm{d}\mathbf{x}, \end{aligned}$$

for $u, u_0 \in H^1(\Omega)$, $0 < \varepsilon, k_\varepsilon \ll 1$, $k_\varepsilon = o(\varepsilon)$, $s \in H^2(\Omega)$, and $\psi(t) := (1-t^2)^2$. So, the edge length term with coefficient α now uses a different (double-well) potential than in the Ambrosio-Tortorelli functional.

1.8 Functionals for Chromaticity and Brightness

In the next two chapters, we are mainly going to study the above approximations to the Mumford-Shah and Mumford-Shah-Euler functionals for unit vector fields. However, in order to actually process colour images, chromaticity and brightness need to be accounted for, so the Mumford-Shah and Mumford-Shah-Euler functionals have to be extended. We assume that chromaticity and brightness share a common edge set, but should otherwise be allowed to evolve separately. This will mean, in particular, that we can apply more smoothing to the chromaticity component or require more fidelity from the brightness component, and the joint edge set will still prohibit "bleeding" of colour information across edges, even if such edges are only discernible in the brightness component.

So, we end up with the following modified Ambrosio-Tortorelli approximation to the Mumford-Shah functional for chromaticity $\mathbf{u} \in H^1(\Omega, \mathbb{S}^{m-1})$ and brightness $u \in H^1(\Omega)$:

$$\begin{aligned} E_\varepsilon(\mathbf{u},u,s) := \frac{\gamma}{2}\int_\Omega \left(s^2 + k_\varepsilon\right)|\nabla \mathbf{u}|^2 \,\mathrm{d}\mathbf{x} + \frac{\gamma_1}{2}\int_\Omega \left(s^2 + k_\varepsilon\right)|\nabla u|^2 \,\mathrm{d}\mathbf{x} + \\ + \frac{\lambda}{2}\int_\Omega |\mathbf{u}-\mathbf{g}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda_1}{2}\int_\Omega |u-g|^2 \,\mathrm{d}\mathbf{x} \\ + \alpha \int_\Omega \left(\varepsilon\,|\nabla s|^2 + \frac{1}{4\varepsilon}(1-s)^2\right)\mathrm{d}\mathbf{x}, \end{aligned}$$

for $\mathbf{u}, \mathbf{g} \in H^1(\Omega, \mathbb{S}^{m-1})$, $u, g \in H^1(\Omega)$, $s \in H^1(\Omega, [0,1])$, $0 < \varepsilon, k_\varepsilon \ll 1$, and $k_\varepsilon = o(\varepsilon)$.

And a corresponding modified approximation to the Mumford-Shah-Euler functional for chromaticity and brightness is

$$\begin{aligned} E_\varepsilon(\mathbf{u}, u, s) :=& \frac{\gamma}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \, \mathrm{d}\mathbf{x} + \frac{\gamma_1}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla u|^2 \, \mathrm{d}\mathbf{x} \\ &+ \frac{\lambda}{2} \int_B |\mathbf{u} - \mathbf{u}_0|^2 \, \mathrm{d}\mathbf{x} + \frac{\lambda_1}{2} \int_B (u - u_0)^2 \, \mathrm{d}\mathbf{x} \\ &+ \alpha \int_\Omega \left(\frac{\varepsilon}{2} |\nabla s|^2 + \frac{1}{\varepsilon} \psi(s) \right) \mathrm{d}\mathbf{x} + \frac{\beta}{\varepsilon} \int_\Omega \left(\frac{1}{\varepsilon} \psi'(s) - \varepsilon \Delta s \right)^2 \mathrm{d}\mathbf{x}, \end{aligned}$$

for $\mathbf{u}, \mathbf{u}_0 \in H^1(\Omega, \mathbb{S}^{m-1})$, $u, u_0 \in H^1(\Omega)$, $s \in H^2(\Omega)$, $0 < \varepsilon, k_\varepsilon \ll 1$, and $k_\varepsilon = o(\varepsilon)$.

Chapter 2

The Mumford-Shah Functional

In this chapter, we propose, analyse, and compare two finite element based numerical approximation schemes for minimising the Mumford-Shah functional for unit vector fields, both analytically and computationally. The first scheme uses a projection strategy, the second a penalisation strategy, to enforce and approximate the sphere constraint, respectively.

Both schemes are then applied to the segmentation of colour images using the chromaticity and brightness colour model.

Numerical results turn out to be more convincing with the projection strategy that enforces the sphere constraint exactly, while the penalisation strategy, that only approximates it, allows for a better convergence result.

2.1 Introduction

For $\Omega \subset \mathbb{R}^d$, and γ, α, λ positive constants, we are interested in numerically minimising the following weak version of the Mumford-Shah energy functional:

$$G(\mathbf{u}) := \frac{\gamma}{2} \int_\Omega |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \alpha \mathcal{H}^{d-1}(S_\mathbf{u}) + \frac{\lambda}{2} \int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x}, \tag{2.1.1}$$

with $\mathbf{u}, \mathbf{g} \in GSBV(\Omega, \mathbb{R}^m)$, and $|\mathbf{u}|^2 = 1$ a.e. (see Section 2.2 for definitions). This is a prototype problem for studying interesting effects with applications in image processing (see e.g. [86, 88, 14, 20, 31, 101, 13]), and liquid crystal theory (see e.g. [75, 78, 38, 102, 1, 12, 26]). Below, we give a short overview over these two applications.

We are sometimes going to refer to functional (2.1.1) as the "Mumford-Shah" functional. It is, in fact, a version (for sphere-valued functions) of a functional proposed by De Giorgi, Carriero, and Leaci in [47] (for scalar functions) as a weak formulation of the original functional proposed by Mumford and Shah in [86] for greyscale image segmentation,

$$E(u, K) := \frac{\gamma}{2} \int_{\Omega \setminus K} |\nabla u|^2 \mathrm{d}\mathbf{x} + \alpha \mathcal{H}^{d-1}(K) + \frac{\lambda}{2} \int_\Omega (u - g)^2 \mathrm{d}\mathbf{x}, \tag{2.1.2}$$

with $g \in L^2(\Omega)$, which is to be minimised for all closed sets $K \subset \Omega$, and functions $u \in H^1(\Omega \setminus K)$. It is shown in [47] that the two problems are essentially equivalent.

The goal of image segmentation is to partition images into meaningful regions, which is often done by finding the edges which bound these regions, and which are in our case identified with the set K. The first term in (2.1.2) ensures smoothness of u outside of K, the second one ensures that there are not too many edges, and the last term ensures that the segmented image u does not deviate too much from the original one g.

A more concrete motivation for studying functional (2.1.1), therefore is *colour image segmentation* in the *Chromaticity and Brightness (CB)* colour model, where the chromaticity (colour information) is represented by an $\mathbb{S}^{m-1}$-valued function (usually $m = 3$) on the image domain Ω. The brightness, represented by a function $b : \Omega \to [0, 1]$, can be separately treated just like a greyscale image. It has been proposed that this model is well-suited for colour image processing, see also Chapter 1. Osher and Vese [88] studied p-harmonic flows to the sphere ($p \geq 1$, in particular $p \in \{1, 2\}$), and applied them to image chromaticity, for example; other sources include [31, 101, 13] and references therein.

The name *free discontinuity problems* was introduced by De Giorgi in [43] for variational problems like (2.1.2), which consist of minimising a functional with volume and surface terms, depending on a closed set K and a function u (usually smooth outside K). Other early sources include [46, 45]. Weak formulations like (2.1.1) allow to prove existence of solutions (see [47] for the scalar, and [27] for the sphere-valued case), but still require the computation of geometric properties of the unknown set of discontinuity boundaries.

Therefore, Ambrosio and Tortorelli introduced an elliptic approximation in [4, 5], whose vectorial version, if defined for sphere-valued functions, is to minimise

$$\begin{aligned} G_\varepsilon(\mathbf{u}, s) := \frac{\gamma}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \, \mathrm{d}\mathbf{x} + \alpha \int_\Omega \left(\varepsilon \, |\nabla s|^2 + \frac{1}{4\varepsilon}(1 - s)^2 \right) \mathrm{d}\mathbf{x} \\ + \frac{\lambda}{2} \int_\Omega |\mathbf{u} - \mathbf{g}|^2 \, \mathrm{d}\mathbf{x} \end{aligned} \tag{2.1.3}$$

for $\mathbf{u}, \mathbf{g} \in H^1(\Omega, \mathbb{S}^{m-1})$, $s \in H^1(\Omega, [0, 1])$, $0 < \varepsilon, k_\varepsilon \ll 1$, and $k_\varepsilon = o(\varepsilon)$. Here, s is a *phase function* approximating $1 - \chi_K$ by penalisation of phase transitions. Ambrosio and Tortorelli showed Γ-convergence of $G_\varepsilon(\mathbf{u}, s)$ to $G(\mathbf{u})$ in L^2 in the scalar ([4, 5]), as well as the $\mathbb{S}^{m-1}$-valued case ([5]) for $\varepsilon \to 0$.

Bellettini and Coscia carried out a finite element approximation of the Mumford-Shah functional in the scalar case, based on this elliptic approximation in [14]. They showed that their approximation $G_{\varepsilon,h} : V^h(\Omega) \times V^h(\Omega, [0, 1]) \to \overline{\mathbb{R}}$ is Γ-convergent to $G : H^1(\Omega) \times H^1(\Omega) \to \overline{\mathbb{R}}$ provided that the mesh size fulfils $h = o(k_\varepsilon)$, and that S_u is piecewise C^2. Here, $V^h(\Omega)$ is the continuous, piecewise affine finite element space. Using the approximation result in [49], Bourdin in [20] showed that S_u does not need to be assumed piecewise C^2; and he proposed an algorithm for actual computations — without providing a proof for its convergence, though. The problem here is that the two variables u and s appear strongly coupled in the energy and in the corresponding gradient flow.

As an alternative to the above phase-field approximation of the Mumford-Shah functional, Braides and Dal Maso proposed a non-local approximation approach in [23], on which Cortesani based a Γ-convergent, vector-valued finite element approximation in [40].

A different motivation for (2.1.3) comes from the theory of *nematic liquid crystals*. In order to overcome mathematical difficulties in showing existence and regularity of energy

minimising static configurations in the Oseen-Frank model, Lin in [75] adapts Ericksen's energy, which he simplifies to (see [75, equation (3.12)])

$$\int_\Omega \frac{1}{2}s^2|\nabla \mathbf{n}|^2 + |\nabla s|^2 + W_0(s)\,\mathrm{d}\mathbf{x}$$

with variable degree of orientation $s \in [-1/2, 1]$ (in experiments, often $s \geq 0$), and director $\mathbf{n}$, $|\mathbf{n}| = 1$ a.e. The strong similarities of this energy to functional (2.1.3) lets us hope that our analysis may be of use to this application, too.

The overall goal of the present work is to construct and analyse convergent discretisations for a prototype problem with several non-convexities; namely, we consider a non-convex functional (the Mumford-Shah functional) with a non-convex constraint (the sphere constraint), as an extension to existing work on convex functionals (in particular harmonic maps) with non-convex constraints, which have been intensely studied (see e.g. [1, 10, 12] and references therein). In particular, we deal with discretisations of the sphere constraint, which we account for using a projection and a penalisation strategy. The former turns out to deliver more convincing computational results, while the latter is analytically more satisfactory.

Below, we give a short overview over the two methods for the approximation of (2.1.1) that we shall present in Sections 3 – 7 of this paper, where we in particular discuss relevant stability properties of computed approximations, such as

- energy decay property for splitting schemes related to (2.1.3),
- the validity of a discrete or penalised sphere constraint for approximations of $\mathbf{u}$, and
- non-negativity and upper bounds for approximations of the phase field function s.

2.1.1 Splitting & Projection Strategy

The problem of coupled variables is addressed through an iterative splitting strategy; i.e., in every step of the iteration the energy is first minimised with respect to the first variable while keeping the second variable fixed, and then minimised with respect to the second variable while keeping the first one fixed. A special projection idea as proposed by Alouges in [1] is used to enforce the sphere constraint. We propose a first-order finite element discretisation, which preserves the sphere constraint exactly at nodal points. The resulting discrete algorithm is simple, results in only linear equations to be solved in every step of the iteration, and can be proved to be energy decreasing. Identifying limits, however has to remain an open problem.

2.1.2 Penalisation & Splitting Strategy

This method again uses a splitting strategy, but the sphere constraint is now approximated by penalisation; i.e., we add a Ginzburg-Landau term $\frac{1}{4\delta_\varepsilon}\int_\Omega \left(|\mathbf{u}|^2 - 1\right)^2 \mathrm{d}\mathbf{x}$ $(0 < \delta \ll 1)$ to the energy (2.1.3). We show that for proper scales of δ_ε in terms of ε, this does not affect Γ-convergence. Furthermore, we propose a first-order finite element algorithm

based on this splitting and penalisation strategy, and show its convergence for vanishing discretisation parameters.

In Section 2.8, comparative computational experiments for the "Penalisation & Splitting" and the "Splitting & Projection" methods are presented, which address in particular

(1) the effect of perturbing the sphere constraint throughout minimisation, as well as proper scalings of regularisation and numerical parameters;

(2) the accuracy of zero sets of s in the course of minimisation; and

(3) comparative numerical studies to relate the CB and RGB models in colour image segmentation, for which we extend the functional (2.1.3) to accommodate both the chromaticity and brightness components.

Finally, Section 2.9 contains a summarising conclusion.

2.2 Preliminaries

We shall use c and C as generic non-negative constants. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $\langle \mathbf{x}, \mathbf{y} \rangle$ or $\mathbf{x} \cdot \mathbf{y}$ will denote their standard scalar product, and $|\mathbf{x}|$ the Euclidean norm of $\mathbf{x}$. For a set S, $|S|$ or $\mathcal{L}^d(S)$ denotes its Lebesgue measure of dimension d, $\mathcal{H}^d(S)$ its Hausdorff measure. The L^2 scalar product and norm will be denoted by $(\cdot, \cdot)$ and $\|\cdot\|$, respectively, and $\mathbb{S}^{m-1}$ will be the unit sphere in $\mathbb{R}^m$. For $a, b \in \mathbb{R}$, let $a \wedge b := \min\{a, b\}$, and $a \vee b := \max\{a, b\}$.

By $A : B$ for $A, B \in \mathbb{R}^{m \times m}$ we shall denote the dyadic product; i.e., $A : B := \sum_{i,j=1}^m a_{ij} b_{ij}$ for $A = (a_{ij})$, $B = (b_{ij})$. Let $|A|$ denote the Frobenius norm of A; i.e., $|A|^2 := \sum_{i,j=1}^m |a_{ij}|^2$. For two vectors $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^m$, let $\mathbf{a} \otimes \mathbf{b} := M$ denote the matrix with entries $m_{ij} := \mathbf{a}_i \mathbf{b}_j$.

We use capital letters for finite element functions and boldface for vectors or vector-valued functions.

2.2.1 Functions of Bounded Variation and Γ-Convergence

We summarise some definitions and results on functions of bounded variation and Γ-convergence. Sources are e.g. [3, 66, 59, 42, 21, 22, 29].

2.2.1.1 BV, SBV, and GSBV Functions

Let $\Omega \subset \mathbb{R}^d$ be a bounded open set, $\mathbf{u} : \Omega \to \mathbb{R}^m$ a measurable function, $S := \mathbb{R}^m \cup \{\infty\}$, and $\mathbf{x} \in \Omega$ be fixed. We call $\mathbf{z} \in S$ the *approximate limit* of $\mathbf{u}$ at $\mathbf{x}$, or $\mathbf{z} = \operatorname{ap} - \lim_{\mathbf{y} \to \mathbf{x}} \mathbf{u}(\mathbf{y})$, if for every neighbourhood U of $\mathbf{z} \in S$ we have

$$\lim_{\varrho \to \infty} \frac{1}{\varrho^n} |\{\mathbf{y} \in \Omega : |\mathbf{y} - \mathbf{x}| < \varrho, \mathbf{u}(\mathbf{y}) \notin U\}| = 0.$$

If $\mathbf{z} \in \mathbb{R}^m$, we call $\mathbf{x}$ a *Lebesgue point* of $\mathbf{u}$, and we denote by $S_{\mathbf{u}}$ the complement of the set of Lebesgue points of $\mathbf{u}$ (*approximate discontinuity set*). Since $|S_{\mathbf{u}}|$ is known to be

zero, $\mathbf{u} = \widetilde{\mathbf{u}}$ a.e. for

$$\widetilde{\mathbf{u}}(\mathbf{x}) := \operatorname{ap} - \lim_{\substack{\mathbf{y}\to\mathbf{x}\\ \mathbf{y}\in\Omega}} \mathbf{u}(\mathbf{y}).$$

Let $\mathbf{x} \in \Omega \setminus S_{\mathbf{u}}$ such that $\widetilde{\mathbf{u}}(\mathbf{x}) \neq \infty$. If there exists $L \in \mathbb{R}^{d\times m}$ such that

$$\operatorname{ap} - \lim_{\substack{\mathbf{y}\to\mathbf{x}\\ \mathbf{y}\in\Omega}} \frac{|\mathbf{u}(\mathbf{y}) - \widetilde{\mathbf{u}}(\mathbf{x}) - L(\mathbf{y}-\mathbf{x})|}{|\mathbf{y}-\mathbf{x}|} = 0,$$

we call $\mathbf{u}$ *approximately differentiable* in $\mathbf{x}$, and $\nabla\mathbf{u}(\mathbf{x}) := L$ the (uniquely determined) *approximate gradient* of $\mathbf{u}$ in $\mathbf{x}$. A function $\mathbf{u} \in L^1(\Omega, \mathbb{R}^m)$ is called a *function of bounded variation* in Ω, or $\mathbf{u} \in BV(\Omega, \mathbb{R}^m)$, if its distributional derivative $D\mathbf{u}$ is representable by a measure with finite total variation $|D\mathbf{u}|(\Omega)$; i.e., if

$$\sum_{\alpha=1}^{m} \int_\Omega u^\alpha \operatorname{div}(\varphi^\alpha)\, \mathrm{d}\mathbf{x} = -\sum_{\alpha=1}^{m}\sum_{i=1}^{d} \int_\Omega \varphi_i^\alpha \mathrm{d}D_i u^\alpha \qquad \forall\, \varphi \in C_c^1\big(\Omega, \mathbb{R}^{m\times d}\big),$$

with $D\mathbf{u}$ an $\mathbb{R}^{d\times m}$ valued matrix of measures $D_i u^\alpha$, and $\mathbf{u} = (u_1, \ldots, u_m)$. Defining

$$\|\mathbf{u}\|_{BV(\Omega,\mathbb{R}^m)} := \|\mathbf{u}\|_{L^1(\Omega,\mathbb{R}^m)} + |D\mathbf{u}|(\Omega),$$

makes $BV(\Omega, \mathbb{R}^m)$ a Banach space.

If $\{\mathbf{u}_j\} \subset BV(\Omega, \mathbb{R}^m)$ with $\sup_j \|\mathbf{u}_j\|_{BV(\Omega,\mathbb{R}^m)} < +\infty$, then there exist a subsequence $\{\mathbf{u}_{j_k}\}$ and a function $\mathbf{u} \in BV(\Omega, \mathbb{R}^m)$ such that $\mathbf{u}_{j_k} \to \mathbf{u}$ in $L^1(\Omega, \mathbb{R}^m)$, and $D\mathbf{u}_{j_k} \to D\mathbf{u}$ weakly-$*$ in the sense of measures.

Also, for $\mathbf{u} \in BV(\Omega, \mathbb{R}^m)$, $S_{\mathbf{u}}$ is countably $\mathcal{H}^{d-1}$-rectifiable; i.e.,

$$S_{\mathbf{u}} = N \cup \bigcup_{i\in\mathbb{N}} K_i,$$

where $\mathcal{H}^{d-1}(N) = 0$, and each K_i is a compact subset of a C^1 manifold. So, for $\mathcal{H}^{d-1}$-a.e. $\mathbf{y} \in S_{\mathbf{u}}$ we can define an *exterior unit normal* $\boldsymbol{\nu}_{\mathbf{u}}$ and *outer* and *inner traces* of $\mathbf{u}$ on $S_{\mathbf{u}}$ by

$$\mathbf{u}^\pm(\mathbf{x}) := \operatorname{ap} - \lim_{\substack{\mathbf{y}\to\mathbf{x}\\ \mathbf{y}\in\pi^\pm(\mathbf{x},\boldsymbol{\nu}_{\mathbf{u}}(\mathbf{x}))}} \mathbf{u}(\mathbf{y}),$$

with $\pi^\pm(\mathbf{x}, \boldsymbol{\nu}_{\mathbf{u}}(\mathbf{x})) := \big\{\mathbf{y} \in \mathbb{R}^d : \pm\langle \mathbf{y}-\mathbf{x}, \boldsymbol{\nu}_{\mathbf{u}}(\mathbf{x})\rangle > 0\big\}$. A point $\mathbf{x} \in \Omega$ is called a *jump point* of $\mathbf{u}$, $\mathbf{x} \in J_{\mathbf{u}}$, if there exists $\boldsymbol{\nu} \in \mathbb{S}^{d-1}$, such that

$$\operatorname{ap} - \lim_{\substack{\mathbf{y}\to\mathbf{x}\\ \mathbf{y}\in\pi^-(\mathbf{x},\boldsymbol{\nu})}} \mathbf{u}(\mathbf{y}) \neq \operatorname{ap} - \lim_{\substack{\mathbf{y}\to\mathbf{x}\\ \mathbf{y}\in\pi^+(\mathbf{x},\boldsymbol{\nu})}} \mathbf{u}(\mathbf{y}).$$

It is known that $J_{\mathbf{u}} \subseteq S_{\mathbf{u}}$, and $\mathcal{H}^{d-1}(S_{\mathbf{u}} \setminus J_{\mathbf{u}}) = 0$.

If we decompose $D\mathbf{u}$ into an absolutely continuous part $D^a\mathbf{u}$ and a singular part $D^s\mathbf{u}$, both with respect to the Lebesgue measure $\mathcal{L}^d$, $D\mathbf{u} = D^a\mathbf{u} + D^s\mathbf{u}$, the density of $D^a\mathbf{u}$ with respect to $\mathcal{L}^d$ coincides with the approximate gradient $\nabla\mathbf{u}$ $\mathcal{L}^d$-a.e. The restriction $D^j\mathbf{u}$ of $D^s\mathbf{u}$ to $S_{\mathbf{u}}$ is called *jump part* of $D\mathbf{u}$, the restriction $D^c\mathbf{u}$ of $D^s\mathbf{u}$ to $\Omega \setminus S_{\mathbf{u}}$ it called *Cantor part*. So,

$$D\mathbf{u} = D^a\mathbf{u} + D^j\mathbf{u} + D^c\mathbf{u}.$$

It is known that $D^j\mathbf{u} = (\mathbf{u}^+ - \mathbf{u}^-) \otimes \boldsymbol{\nu}_\mathbf{u} \mathcal{H}^{d-1} \lfloor S_\mathbf{u}$.

A function $\mathbf{u} \in BV(\Omega, \mathbb{R}^m)$ is called a *special function of bounded variation* in Ω, $\mathbf{u} \in SBV(\Omega, \mathbb{R}^m)$, if $D^c\mathbf{u} = 0$. We call $\mathbf{u} \in BV(\Omega, \mathbb{R}^m)$ a *generalised special function of bounded variation*, $\mathbf{u} \in GSBV(\Omega, \mathbb{R}^m)$, if $\mathbf{g}(\mathbf{u}) \in SBV(\Omega, \mathbb{R}^m)$ for every $\mathbf{g} \in C^1(\mathbb{R}^m)$ such that $\nabla \mathbf{g}$ has compact support. For $1 < p < +\infty$, let

$$(G)SBV^p(\Omega, \mathbb{R}^m) := \left\{ \mathbf{u} \in (G)SBV(\Omega, \mathbb{R}^m) : \mathcal{H}^{d-1}(J_\mathbf{u}) < +\infty,\ \nabla \mathbf{u} \in L^p\left(\Omega, \mathbb{R}^{d \times m}\right) \right\}.$$

We remark that $W^{1,1}(\Omega, \mathbb{R}^m) \subsetneq BV(\Omega, \mathbb{R}^m)$, that $\mathbf{u} \in SBV(\Omega, R^m)$ implies $\mathbf{u} \in W^{1,1}\left(\Omega \setminus \overline{S_\mathbf{u}}, \mathbb{R}^m\right)$, and that $SBV(\Omega, \mathbb{R}^m) \cap L^\infty(\Omega, \mathbb{R}^m) = GSBV(\Omega, \mathbb{R}^m) \cap L^\infty(\Omega, \mathbb{R}^m)$.

2.2.1.2 Γ-Convergence

Let X be a separable Banach space with a topology τ and let $F_\varepsilon : X \to \overline{\mathbb{R}}$ be a sequence of functionals. We say F_ε Γ-*converges* to F in the topology τ, or $F = \Gamma - \lim_{\varepsilon \to 0} F_\varepsilon$, if the following two conditions hold:

(1) For every $x \in X$ and for every sequence $\{x_\varepsilon\} \subset X$ τ-converging to $x \in X$,

$$F(x) \leq \liminf_{\varepsilon \to 0} F_\varepsilon(x_\varepsilon),$$

(2) For every $x \in X$ there exists a sequence $\{x_\varepsilon\} \subset X$ (*recovery sequence*) τ-converging to $x \in X$, such that

$$F(x) \geq \limsup_{\varepsilon \to 0} F_\varepsilon(x_\varepsilon).$$

Lemma 2.2.1. *Let $F_\varepsilon, F : X \to \overline{\mathbb{R}}$, with $\Gamma - \lim_{\varepsilon \to 0} F_\varepsilon = F$. Then*

(1) *F is lower semicontinuous on X.*

(2) *$F + G = \Gamma - \lim (F_\varepsilon + G)$ for all continuous $G : X \to \mathbb{R}$.*

(3) *Let $\{\mathbf{u}_\varepsilon\} \subset X$ be such that*

$$\lim_{\varepsilon \to 0^+} \left(F_\varepsilon(\mathbf{u}_\varepsilon) - \inf_X F_\varepsilon \right) = 0,$$

then every cluster point $\mathbf{u}$ of $\{\mathbf{u}_\varepsilon\}$ minimises F over X, and

$$\lim_{\varepsilon \to 0^+} \inf_X F_\varepsilon = \min_X F = F(\mathbf{u}).$$

Here are some connections between Γ-convergence and pointwise convergence:

- If F_ε converges uniformly to F, then F_ε Γ-converges to F.
- If F_ε is decreasing and converges pointwise to F, then F_ε Γ-converges to RF, the lower semicontinuous envelope of F.

2.3 Continuous Algorithm for Splitting & Projection

2.3.1 Algorithm

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. For $\mathbf{u}, \mathbf{g} \in H^1(\Omega, \mathbb{S}^{m-1})$, $s \in H^1(\Omega, [0,1])$, and $0 < \varepsilon, k_\varepsilon \ll 1$ we want to minimise the following vector valued Ambrosio-Tortorelli energy:

$$\begin{aligned} AT_\varepsilon(\mathbf{u}, s) &:= \frac{\gamma}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \, \mathrm{d}\mathbf{x} + \frac{\lambda}{2} \int_\Omega |\mathbf{u} - \mathbf{g}|^2 \, \mathrm{d}\mathbf{x} \\ &\quad + \alpha \int_\Omega \varepsilon \, |\nabla s|^2 + \frac{1}{4\varepsilon}(1-s)^2 \mathrm{d}\mathbf{x}. \end{aligned} \tag{2.3.1}$$

In this section, we shall always assume $\gamma, \alpha, \varepsilon, k_\varepsilon$ to be fixed and positive, and $\lambda \geq 0$. Furthermore, let

$$\begin{aligned} K_{\mathbf{u}} &:= \left\{ \mathbf{w} \in H^1(\Omega, \mathbb{R}^m) : \mathbf{w} \cdot \mathbf{u} = 0 \text{ a.e.} \right\}, \text{ and} \\ \widetilde{AT}_\varepsilon(\mathbf{u}, s) &:= \frac{1}{2} \int_\Omega \gamma \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 + \lambda |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x}. \end{aligned}$$

Algorithm 2.3.1. (1) *Let* $(\mathbf{u}_0, s_0) \in H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega, [0,1])$ *be given.*

(2) *For* $n = 0, \ldots$ *until convergence*

(a) Minimise $\widetilde{AT}_\varepsilon(\mathbf{u}_n - \mathbf{w}, s_n)$ *for all* $\mathbf{w} \in K_{\mathbf{u}_n}$*, and call the solution* $\mathbf{w}_n$*.*

(b) Set $\mathbf{u}_{n+1} := (\mathbf{u}_n - \mathbf{w}_n) \,/\, |\mathbf{u}_n - \mathbf{w}_n|$*.*

(c) Minimise $AT_\varepsilon(\mathbf{u}_{n+1}, s)$ *for all* $s \in H^1(\Omega)$*, and call the solution* s_{n+1}*.*

Following Alouges in [1], we have transformed the nonlinear, non-convex constraint $|\mathbf{u}|^2 = 1$ a.e. into a linear one, namely $\mathbf{u} \cdot \mathbf{w} = 0$ a.e.

2.3.2 Analysis

The next proposition ensures that the solution $\mathbf{u}_n - \mathbf{w}$ found in each iteration n of Step 2a of Algorithm 2.3.1 has at least length 1. This is crucial for proving that the projection in Step 2b, and ultimately the whole Algorithm 2.3.1, is energy decreasing, as we show in Proposition 2.3.5.

Proposition 2.3.2. *Let* $\mathbf{u} \in H^1(\Omega, \mathbb{S}^{m-1})$*. If for* $\mathbf{w} \in K_{\mathbf{u}}$ *we set* $\mathbf{v} := \mathbf{u} - \mathbf{w}$*, then*

$$|\mathbf{v}(\mathbf{x})|^2 = |\mathbf{u}(\mathbf{x}) - \mathbf{w}(\mathbf{x})|^2 = 1 + |\mathbf{w}(\mathbf{x})|^2 \geq 1 \text{ a.e.}$$

Proposition 2.3.3. *Given* $\mathbf{u}, \mathbf{g} \in H^1(\Omega, \mathbb{S}^{m-1})$*,* $s \in H^1(\Omega)$*, and parameters* $\varepsilon, k_\varepsilon > 0$*, the problem*

$$\text{Minimise } \widetilde{AT}_\varepsilon(\mathbf{u} - \mathbf{w}, s) \quad \text{for } \mathbf{w} \in K_{\mathbf{u}} \tag{2.3.2}$$

has a unique solution.

Proof. $K_{\mathbf{u}}$ is a non-empty subspace of $H^1(\Omega, \mathbb{R}^m)$. It is closed in the strong topology of $H^1(\Omega, \mathbb{R}^m)$ and $AT_\varepsilon(\mathbf{v}, s)$ is uniformly elliptic in $\mathbf{v}$. Therefore there is a unique solution to our restricted optimisation problem.

Note: The mapping $\mathbf{u} \mapsto \mathbf{w}(\mathbf{u})$ is the projection of $\mathbf{u}$ onto $K_{\mathbf{u}}$ with respect to the usual distance in $H^1(\Omega, \mathbb{R}^m)$. And since $\mathbf{0} \in K_{\mathbf{u}}$, we have $AT_\varepsilon(\mathbf{u} - \mathbf{w}, s) \leq AT_\varepsilon(\mathbf{u}, s)$. □

Lemma 2.3.4. *The tuple $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega, [0,1])$ is a stationary point of $AT_\varepsilon(\cdot, \cdot)$ if and only if*

$$\gamma \left(\left(s^2 + k_\varepsilon \right) \nabla \mathbf{u}, \nabla \boldsymbol{\varphi} \right) = \lambda(\mathbf{g}, \boldsymbol{\varphi}) \tag{2.3.3}$$

for all $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$ such that $\boldsymbol{\varphi}(\mathbf{x}) \in T_{\mathbf{u}(\mathbf{x})}\mathbb{S}^{m-1}$ (the tangent space of $\mathbb{S}^{m-1}$ at $\mathbf{u}(\mathbf{x})$), and

$$2\alpha\varepsilon(\nabla s, \nabla \varphi) + \left(\left(\gamma |\nabla \mathbf{u}|^2 + \frac{\alpha}{2\varepsilon} \right) s, \varphi \right) = \left(\frac{\alpha}{2\varepsilon}, \varphi \right) \tag{2.3.4}$$

for all $\varphi \in H^1(\Omega) \cap L^\infty(\Omega)$.

Proof. Note $\mathbf{u} \cdot \boldsymbol{\varphi} = 0$ a.e. and derive the first variation of (2.3.1) with respect to $\mathbf{u}$ and s, c.f. [99] and [25, Proposition 1]. □

Most of the following Proposition is taken from [2].

Proposition 2.3.5. *If $\mathbf{v} \in H^1(\Omega, \mathbb{R}^m)$ verifies $|\mathbf{v}| \geq 1$ a.e., then $\mathbf{v}/|\mathbf{v}|$ belongs to $H^1(\Omega, \mathbb{S}^{m-1})$ and*

$$\left| \nabla \left(\frac{\mathbf{v}}{|\mathbf{v}|} \right) \right|^2 \leq |\nabla \mathbf{v}|^2 \text{ a.e.,}$$

which, for given $s \in H^1(\Omega)$ and $\varepsilon, k_\varepsilon > 0$ as above, leads to

$$AT_\varepsilon \left(\frac{\mathbf{v}}{|\mathbf{v}|}, s \right) \leq AT_\varepsilon(\mathbf{v}, s) \,,$$

and $\widetilde{AT}_\varepsilon(\mathbf{v}/|\mathbf{v}|, s) \leq \widetilde{AT}_\varepsilon(\mathbf{v}, s)$.

Proof. By definition, $\mathbf{v}/|\mathbf{v}| \in H^1(\Omega, \mathbb{S}^{m-1})$. And

$$\begin{aligned}
|\nabla \mathbf{v}|^2 &= \left| \nabla \left(|\mathbf{v}| \frac{\mathbf{v}}{|\mathbf{v}|} \right) \right|^2 \\
&= \left| (\nabla |\mathbf{v}|) \otimes \frac{\mathbf{v}}{|\mathbf{v}|} + |\mathbf{v}| \nabla \left(\frac{\mathbf{v}}{|\mathbf{v}|} \right) \right|^2 \\
&= |\nabla |\mathbf{v}||^2 + \left| \nabla \left(\frac{\mathbf{v}}{|\mathbf{v}|} \right) \right|^2 |\mathbf{v}|^2 \\
&\geq \left| \nabla \left(\frac{\mathbf{v}}{|\mathbf{v}|} \right) \right|^2 \text{ a.e.,}
\end{aligned}$$

since $|\mathbf{v}| \geq 1$ a.e. and $\left| \frac{\mathbf{v}}{|\mathbf{v}|} \right| = 1$, which implies $\frac{\mathbf{v}}{|\mathbf{v}|} \nabla \left(\frac{\mathbf{v}}{|\mathbf{v}|} \right) = 0$.

So, if we are mapping $\mathbf{v} \mapsto \mathbf{v}/|\mathbf{v}|$, the first term (smoothing term) of (2.3.1) does not increase. The third term (edge-length term) obviously stays the same. And for the second term (fidelity term), we have

$$\left| \frac{\mathbf{v}}{|\mathbf{v}|} - \mathbf{g} \right|^2 \leq |\mathbf{v} - \mathbf{g}|^2 ,$$

by a simple computation. □

Proposition 2.3.6. *There exists a unique $s \in H^1(\Omega, [0,1])$, such that for all $\varphi \in H^1(\Omega)$*

$$a\left(\nabla s, \nabla \varphi\right) + \left(bs, \varphi\right) = c\left(1, \varphi\right), \tag{2.3.5}$$

with $a = 2\alpha\varepsilon$, $b(\mathbf{x}) = \gamma|\nabla \mathbf{u}(\mathbf{x})|^2 + \frac{\alpha}{2\varepsilon}$, and $c = \frac{\alpha}{2\varepsilon}$. Hence, Step 2c of Algorithm 2.3.1 is solvable.

Proof. Existence and uniqueness of a solution are standard (by coercivity and convexity of the energy in ∇s).

To show $0 \leq s \leq 1$ a.e. we apply (2.3.5) to s and $t := s-1$ and test it with $s^- := -s \vee 0$ and $t^+ := t \vee 0$, respectively, obtaining

$$\begin{aligned} 0 &\leq a \int_{\{-s \geq 0\}} \nabla s \nabla(-s) \mathrm{d}\mathbf{x} + \int_{\{-s \geq 0\}} bs(-s) \mathrm{d}\mathbf{x} \\ &= -a \int_{\{-s \geq 0\}} |\nabla s|^2 \, \mathrm{d}\mathbf{x} - \int_{\{-s \geq 0\}} b\,|s|^2 \, \mathrm{d}\mathbf{x}, \end{aligned}$$

which can only be fulfilled if $s \geq 0$ a.e. Analogously, we get

$$0 \geq a \int_{\{t \geq 0\}} |\nabla t|^2 \, \mathrm{d}\mathbf{x} + \int_{\{t \geq 0\}} b\,|t|^2 \, \mathrm{d}\mathbf{x},$$

which can only be true if $t = s - 1 \leq 0$ a.e. □

Remark 2.3.7. *The fact that $0 \leq s \leq 1$ in the proposition above could also be proved by a simple energy argument, as observed in [20]. We use a similar argument in the discrete case in Section 2.4.*

Remark 2.3.8. *Algorithm 2.3.1 is energy-decreasing, since for any applicable n it ensures*

$$\begin{aligned} AT_\varepsilon(\mathbf{u}_{n+1}, s_{n+1}) &\leq AT_\varepsilon(\mathbf{u}_{n+1}, s_n) \\ &= AT_\varepsilon\left(\frac{\mathbf{u}_n - \mathbf{w}_n}{|\mathbf{u}_n - \mathbf{w}_n|}, s_n\right) \\ &\leq AT_\varepsilon(\mathbf{u}_n - \mathbf{w}_n, s_n) \\ &\leq AT_\varepsilon(\mathbf{u}_n, s_n) . \end{aligned}$$

Lemma 2.3.9. *Let $n \geq 0$. Then*

$$\frac{1}{2} \int_\Omega \gamma \left(s_n^2 + k_\varepsilon\right) |\nabla \mathbf{w}_n|^2 + \lambda \, |\mathbf{w}_n|^2 \, \mathrm{d}\mathbf{x} \leq \widetilde{AT}_\varepsilon(\mathbf{u}_n, s_n) - \widetilde{AT}_\varepsilon(\mathbf{u}_{n+1}, s_{n+1}) .$$

Proof. By Proposition 2.3.5, we have

$$\frac{1}{2}I := \widetilde{AT}_\varepsilon(\mathbf{u}_{n+1}, s_{n+1}) - \widetilde{AT}_\varepsilon(\mathbf{u}_n, s_n) \leq \widetilde{AT}_\varepsilon(\mathbf{u}_n - \mathbf{w}_n, s_n) - \widetilde{AT}_\varepsilon(\mathbf{u}_n, s_n)\,.$$

Expanding the above, we get

$$\begin{aligned} I \leq\ & \gamma \int_\Omega \left(s_n^2 + k_\varepsilon\right) \left(|\nabla \mathbf{u}_n|^2 + |\nabla \mathbf{w}_n|^2 - 2\nabla \mathbf{u}_n : \nabla \mathbf{w}_n\right) \mathrm{d}\mathbf{x} \\ & + \lambda \int_\Omega |\mathbf{u}_n|^2 + |\mathbf{w}_n|^2 + |\mathbf{g}|^2 - 2\mathbf{g}\cdot(\mathbf{u}_n - \mathbf{w}_n) - 2\mathbf{u}_n \cdot \mathbf{w}_n \mathrm{d}\mathbf{x} \\ & - \int_\Omega \gamma \left(s_n^2 + k_\varepsilon\right) |\nabla \mathbf{u}_n|^2 + \lambda \left(|\mathbf{u}_n|^2 + |\mathbf{g}|^2 - 2\mathbf{g}\cdot\mathbf{u}_n\right) \mathrm{d}\mathbf{x} \\ & = \int_\Omega \gamma \left(s_n^2 + k_\varepsilon\right) \left(|\nabla \mathbf{w}_n|^2 - 2\nabla \mathbf{u}_n : \nabla \mathbf{w}_n\right) + \lambda \left(|\mathbf{w}_n|^2 + 2\mathbf{w}_n \cdot (\mathbf{g} - \mathbf{u}_n)\right) \mathrm{d}\mathbf{x}. \end{aligned} \tag{2.3.6}$$

Here we deliberately did not use $\mathbf{u}_n \cdot \mathbf{w}_n = 0$ a.e. and $|\mathbf{u}_n|^2 = 1 = |\mathbf{w}_n|^2$ a.e., so we can use the same proof in the discrete case. The variational formulation of problem (2.3.2) is

$$\int_\Omega \gamma \left(s_n^2 + k_\varepsilon\right) \nabla\left(\mathbf{u}_n - \mathbf{w}_n\right) : \nabla \boldsymbol{\varphi} + \lambda \left(\mathbf{u}_n - \mathbf{w}_n - \mathbf{g}\right) \cdot \boldsymbol{\varphi}_n \mathrm{d}\mathbf{x} = 0 \quad \forall\, \boldsymbol{\varphi}_n \in K_{\mathbf{u}_n}.$$

Setting $\boldsymbol{\varphi}_n := \mathbf{w}_n$ leads to

$$\int_\Omega \gamma \left(s_n^2 + k_\varepsilon\right) \nabla \mathbf{u}_n : \nabla \mathbf{w}_n \mathrm{d}\mathbf{x} = \int_\Omega \gamma \left(s_n^2 + k_\varepsilon\right) |\nabla \mathbf{w}_n|^2 + \lambda \left(|\mathbf{w}_n|^2 + \mathbf{w}_n \cdot (\mathbf{g} - \mathbf{u}_n)\right) \mathrm{d}\mathbf{x}.$$

So (2.3.6) becomes

$$I \leq -\int_\Omega \gamma \left(s_n^2 + k_\varepsilon\right) |\nabla \mathbf{w}_n|^2 + \lambda |\mathbf{w}_n|^2 \,\mathrm{d}\mathbf{x},$$

as claimed. □

Proposition 2.3.10. *If $AT_\varepsilon(\mathbf{u}_0, s_0)$ is bounded, Algorithm 2.3.1 converges in the sense that $(\mathbf{u}_n, s_n)$ converges weakly (up to a subsequence) in $H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega)$ to some $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega, [0,1])$. Furthermore, $\mathbf{w}_n \to 0$ strongly in $H^1(\Omega, \mathbb{R}^m)$ (the whole sequence).*

Proof. **Step 1:** *$(\mathbf{u}_n, s_n)$ converges weakly (up to a subsequence) in $H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega)$ to some $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega, [0,1])$.*

By assumption and Remark 2.3.8, the L^2 norms $\|\nabla \mathbf{u}_n\|, \|\mathbf{u}_n\|, \|\nabla s_n\|$, and $\|s_n\|$ remain uniformly bounded. Hence we can extract subsequences $\{\mathbf{u}_n, s_n\}$ (not relabelled), that converge weakly in H^1, strongly in L^2, and a.e. to some $(\mathbf{u}, s)$. Since $\||\mathbf{u}_n| - |\mathbf{u}|\| \leq \|\mathbf{u}_n - \mathbf{u}\| \to 0$, we have $|\mathbf{u}|^2 = 1$ a.e., so $\mathbf{u}$ still belongs to $H^1(\Omega, \mathbb{S}^{m-1})$.

Finally, since $H^1(\Omega)$ is a Hilbert space and $\{\varphi \in H^1(\Omega) : 0 \leq \varphi \leq 1 \text{ a.e}\} \subset H^1(\Omega)$ is a closed, convex set, it is weakly closed. Therefore, by the weak convergence in H^1 of $s_n \rightharpoonup s$, we get $0 \leq s \leq 1$.

Step 2: *$\mathbf{w}_n \to 0$ in $H^1(\Omega, \mathbb{R}^m)$.*

Let $n \in \mathbb{N}$ be fixed. By Lemma 2.3.9, we get

$$\frac{1}{2}\int_\Omega \gamma\left(s_n^2 + k_\varepsilon\right)|\nabla \mathbf{w}_n|^2 + \lambda\,|\mathbf{w}_n|^2\,\mathrm{d}\mathbf{x} \leq \widetilde{AT}_\varepsilon(\mathbf{u}_n, s_n) - \widetilde{AT}_\varepsilon(\mathbf{u}_{n+1}, s_{n+1})\,.$$

Summing this from 0 to N leads to

$$\begin{aligned}\frac{1}{2}\sum_{n=0}^{N}\int_\Omega \gamma\left(s_n^2 + k_\varepsilon\right)|\nabla \mathbf{w}_n|^2 + \lambda\,|\mathbf{w}_n|^2\,\mathrm{d}\mathbf{x} &\leq \widetilde{AT}_\varepsilon(\mathbf{u}_0, s_0) - \widetilde{AT}_\varepsilon(\mathbf{u}_{N+1}, s_{N+1}) \\ &\leq \widetilde{AT}_\varepsilon(\mathbf{u}_0, s_0) < +\infty;\end{aligned}$$

i.e., the series

$$\frac{1}{2}\sum_{n\geq 0}\int_\Omega \gamma\left(s_n^2 + k_\varepsilon\right)|\nabla \mathbf{w}_n|^2 + \lambda\,|\mathbf{w}_n|^2\,\mathrm{d}\mathbf{x}$$

is convergent. Therefore, $0 \leq \gamma k_\varepsilon\,|\nabla \mathbf{w}_n|^2 + \lambda\,|\mathbf{w}_n|^2 \leq \gamma\,(s_n^2 + k_\varepsilon)\,|\nabla \mathbf{w}_n|^2 + \lambda\,|\mathbf{w}_n|^2 \to 0$ a.e. And since $\gamma, \lambda, k_\varepsilon > 0$ are fixed, we get $\mathbf{w}_n \to 0$ strongly in H^1.

□

Remark 2.3.11. *We cannot prove that $(\mathbf{u}, s)$ is a stationary point of $AT_\varepsilon(\cdot,\cdot)$. In particular, the variational formulation of Step 2c of Algorithm 2.3.1 is (c.f. Lemma 2.3.4)*

$$2\alpha\varepsilon\,(\nabla s_n, \nabla\varphi) + \gamma\left(|\nabla \mathbf{u}_{n+1}|^2\,s_n, \varphi\right) + \frac{\alpha}{2\varepsilon}\,(s_n, \varphi) = \frac{\alpha}{2\varepsilon}\,(1, \varphi)$$

for all $\varphi \in H^1(\Omega)$. Identifying limits on a term by term basis would require identifying the limit

$$\lim_{n\to+\infty}\left(|\nabla \mathbf{u}_{n+1}|^2\,s_n, \varphi\right),$$

which so far we have to leave as an open problem.

What is missing for this identification of limits is strong convergence of $\nabla \mathbf{u}_n$ in L^2 (and higher regularity of s_n would be helpful, too).

This missing strong convergence is a fundamental shortcoming also observed in [1, 10] for the simpler case of harmonic maps to the sphere. In fact, we are not aware of any algorithm, even in the harmonic mapping case, that simultaneously gives strong convergence of $\nabla \mathbf{u}_n$ in L^2 and assures the sphere constraint exactly*.*

However, the algorithm converges, decreases the energy, assures the sphere constraint exactly and delivers very convincing computational results (indeed, it is faster and delivers better results than the alternative algorithm described in the sequel, c.f. Section 2.8). Therefore, the next section describes a discrete version of this algorithm with its analysis, which has the same shortcoming (we use the continuous version to describe it, because it is less cluttered).

2.4 Discrete Algorithm for Splitting & Projection

2.4.1 Algorithm

Let $\Omega \subset \mathbb{R}^d$ be a polyhedral Lipschitz domain, and $\mathcal{T}_h$ be a quasi-uniform triangulation of Ω with node set $\mathcal{N}$ and maximal mesh size $h > 0$ (c.f. [24]). The space of globally continuous, piecewise affine finite element functions on $\mathcal{T}_h$ is denoted by $V_h(\Omega) \subseteq H^1(\Omega)$. The nodal basis functions are $\{\varphi_{\mathbf{z}} : \mathbf{z} \in \mathcal{N}\} \subseteq V_h(\Omega)$. Let $V_h(\Omega, \mathbb{R}^m)$ be the finite element space of $\mathbb{R}^m$-valued mappings with basis functions $\{\boldsymbol{\varphi}^i_{\mathbf{z}} : \mathbf{z} \in \mathcal{N}, 1 \le i \le m\}$, with $\boldsymbol{\varphi}^1_{\mathbf{z}} := (\varphi_{\mathbf{z}}, 0, \ldots)^T \in V_h(\Omega, \mathbb{R}^m)$, $\boldsymbol{\varphi}^2_{\mathbf{z}} := (0, \varphi_{\mathbf{z}}, 0, \ldots)^T \in V_h(\Omega, \mathbb{R}^m)$, and so forth. Let $\mathcal{I}_h(\cdot) : C^0(\overline{\Omega}) \to V_h(\Omega)$ be the Lagrange interpolation operator, and $R_h(\cdot) : H^1(\Omega) \to V_h(\Omega)$ the Ritz projection, defined by

$$(\nabla (R_h(\varphi) - \varphi), \nabla V) + (R_h(\varphi) - \varphi, V) = 0 \quad \forall\, V \in V_h(\Omega),$$

and $r_h(\cdot) : L^2(\Omega) \to V_h(\Omega)$ the Clément operator [36] ($\boldsymbol{\mathcal{I}}_h(\cdot)$, $\mathbf{R}_h(\cdot)$, and $\mathbf{r}_h(\cdot)$ in the vector valued case). The latter operator will be needed since it can be applied to non-continuous functions.

The most natural approach to the discrete case would be to work with the original functional $AT_\varepsilon(\cdot,\cdot)$, like in the continuous case. However, it is not clear how to get an L^∞ bound on iterates S_n (or, more specifically, $0 \le S_n \le 1$) in this setting. The straightforward remedy is to use mass lumping in all nonlinear terms; i.e., to use the functional

$$\begin{aligned}&\frac{\gamma}{2}\int_\Omega \left(\mathcal{I}_h\left(S^2\right) + k_\varepsilon\right) |\nabla \mathbf{U}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega |\mathbf{U} - \mathbf{G}|^2 \,\mathrm{d}\mathbf{x}\\ &+\alpha \int_\Omega \varepsilon\, |\nabla S|^2 + \frac{1}{4\varepsilon}\mathcal{I}_h\left((1-S)^2\right) \mathrm{d}\mathbf{x},\end{aligned}$$

see e.g. [25]. This introduces additional error terms to take care of in the convergence analysis, but it would work in our case, too. However, we attempt to avoid mass lumping as far as possible, and we present some ideas to this end. So far, we still have to use lumping in the last term, and our arguments only work for $d \le 2$ (the rest of the analysis works for $d \le 3$), but we hope it will be possible to improve these results. So, for $\mathbf{G} \in V_h(\Omega, \mathbb{R}^m)$, we define

$$\begin{aligned}E_h(\mathbf{U}, S) \;:=\; &\frac{\gamma}{2}\int_\Omega \left(S^2 + k_\varepsilon\right) |\nabla \mathbf{U}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega |\mathbf{U} - \mathbf{G}|^2 \,\mathrm{d}\mathbf{x}\\ &+\alpha \int_\Omega \varepsilon\, |\nabla S|^2 + \frac{1}{4\varepsilon}\mathcal{I}_h\left((1-S)^2\right) \mathrm{d}\mathbf{x},\end{aligned}$$

and

$$\widetilde{E}(\mathbf{U}, S) := \frac{\gamma}{2}\int_\Omega \left(S^2 + k_\varepsilon\right) |\nabla \mathbf{U}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega |\mathbf{U} - \mathbf{G}|^2 \,\mathrm{d}\mathbf{x}.$$

In this section, we shall always assume $\gamma, \alpha, \varepsilon, k_\varepsilon$ to be fixed and positive, $\lambda \ge 0$, and $d \le 2$.

Functions $\mathbf{V} \in V_h(\Omega, \mathbb{R}^m)$ which satisfy the pointwise constraint $|\mathbf{V}| = 1$ are necessarily constant. So it is more reasonable to replace $H^1(\Omega, \mathbb{S}^{m-1})$ by

$$H_h^1(\mathcal{T}_h) := \left\{ \mathbf{V} \in V_h(\Omega, \mathbb{R}^m) : \mathbf{V}(\mathbf{z}) \in \mathbb{S}^{m-1} \ \forall\, \mathbf{z} \in \mathcal{N} \right\}.$$

We set

$$K_h^n := \left\{ \mathbf{W} \in V_h(\Omega, \mathbb{R}^m) : \mathbf{W}(\mathbf{z}) \cdot \mathbf{U}_n(\mathbf{z}) = 0 \ \forall\, \mathbf{z} \in \mathcal{N} \right\},$$

where $\mathbf{U}_n \in H_h^1(\mathcal{T}_h)$ will be the iterates of the fully discrete algorithm.

For calculating a solution, we are motivated by [1] and [10] to propose

Algorithm 2.4.1. *Let a quasi-uniform triangulation $\mathcal{T}_h$ of Ω, starting values $\mathbf{U}_0, S_0$, and parameters $\varepsilon, k_\varepsilon, \varrho > 0$ be given. For $n := 0, \ldots$*

(1) *Minimise $\widetilde{E}(\mathbf{U}_n - \mathbf{W}, S_n)$ for $\mathbf{W} \in K_h^n$; i.e. solve*

$$\gamma \left(\left(S_n^2 + k_\varepsilon \right) \nabla \left(\mathbf{U}_n - \mathbf{W} \right), \nabla \mathbf{V} \right) - \lambda \left(\mathbf{W} + \mathbf{G}, \mathbf{V} \right) = 0, \tag{2.4.1}$$

for all $\mathbf{V} \in K_h^n$, and call the solution $\mathbf{W}_n$.

(2) *If $\|\mathbf{W}_n\|_{H^1(\Omega;\mathbb{R}^m)} \leq \varrho$ set $\mathbf{U} := \mathbf{U}_n$, $\mathbf{W} := \mathbf{W}_n$, $S := S_n$ and stop.*

(3) *Set*

$$\mathbf{U}_{n+1} := \sum_{\mathbf{z} \in \mathcal{N}} \frac{\mathbf{U}_n(\mathbf{z}) - \mathbf{W}_n(\mathbf{z})}{|\mathbf{U}_n(\mathbf{z}) - \mathbf{W}_n(\mathbf{z})|} \boldsymbol{\varphi}_{\mathbf{z}}.$$

(4) *Minimise $E_h(\mathbf{U}_{n+1}, S)$ for all $S \in V_h(\Omega)$; i.e. solve*

$$2\alpha\varepsilon \left(\nabla S, \nabla W \right) + \gamma \left(S \left| \nabla \mathbf{U}_{n+1} \right|^2, W \right) + \frac{\alpha}{2\varepsilon} \left(S - 1, W \right)_h = 0 \tag{2.4.2}$$

for all $W \in V_h(\Omega)$, and call the solution S_{n+1}.

Here $(\varphi, \psi)_h := \int_\Omega \mathcal{I}_h(\varphi\psi)\, \mathrm{d}\mathbf{x}$ for $\varphi, \psi \in C(\overline{\Omega})$.

2.4.2 Analysis

Definition 2.4.2. *Let $\mathcal{T}_h$ be a quasi-uniform triangulation of Ω, and $s \in H^1(\Omega)$ be fixed. $\mathcal{T}_h$ is said to satisfy an* energy decreasing condition (ED) *if*

$$E_h(\mathbf{W}, s) \leq E_h(\mathbf{V}, s)$$

for all $\mathbf{V} \in V_h(\Omega, \mathbb{R}^m)$ fulfilling $|\mathbf{V}(\mathbf{z})| \geq 1$ for $\mathbf{z} \in \mathcal{N}$, and $|\mathbf{V}(\mathbf{z})| = 1$ for $\mathbf{z} \in \mathcal{N} \cap \partial\Omega$. Here $\mathbf{W} \in V_h(\Omega, \mathbb{R}^m)$ is defined by

$$\mathbf{W} := \sum_{\mathbf{z} \in \mathcal{N}} \frac{\mathbf{V}(\mathbf{z})}{|\mathbf{V}(\mathbf{z})|} \varphi_{\mathbf{z}}.$$

We quote two results from [10], which give sufficient conditions for (ED):

Lemma 2.4.3. *Let $\mathcal{T}_h$ be a regular triangulation of Ω and $\int_\Omega \nabla\varphi_{\mathbf{z}} \cdot \nabla\varphi_{\mathbf{y}} \mathrm{d}\mathbf{x} \leq 0$ for all $\mathbf{z} \in \mathcal{N} \setminus \partial\Omega$ and $\mathbf{y} \in \mathcal{N} \setminus \{\mathbf{z}\}$. Then $\mathcal{T}_h$ satisfies (ED).*

Corollary 2.4.4. (1) *Let $d = 2$. Given neighbouring nodes $\mathbf{z} \in \mathcal{N}\setminus\partial\Omega$ and $\mathbf{y} \in \mathcal{N}\setminus\{\mathbf{z}\}$, let $T_1, T_2 \in \mathcal{T}_h$ be such that $T_1 \cap T_2$ equals the interior edge connecting $\mathbf{z}$ and $\mathbf{y}$. Let $\alpha^1_{\mathbf{zy}}, \alpha^2_{\mathbf{zy}}$ be the angles of T_1 and T_2, respectively, opposite to the edge connecting $\mathbf{z}$ and $\mathbf{y}$. Then $\alpha^1_{\mathbf{zy}} + \alpha^2_{\mathbf{zy}} \leq \pi$ is sufficient for (ED).*

(2) *Let $d = 3$, and $\mathbf{z} \in \mathcal{N} \setminus \partial\Omega$ and $\mathbf{y} \in \mathcal{N} \setminus \{\mathbf{z}\}$ be nodes of the same Element $T \subset \mathcal{T}_h$. Given any $T \in \mathcal{T}_h$ such that $\mathbf{z}, \mathbf{y} \in T$, let $\alpha_{\mathbf{zy}T}$ be the angle between the two faces $F^1_{\mathbf{zy}}, F^2_{\mathbf{zy}} \subset \partial T$, which do not contain both $\mathbf{z}$ and $\mathbf{y}$. Then $\alpha_{\mathbf{zy}T} \leq \pi/2$ for all $T \ni \mathbf{z}, \mathbf{y}$ is sufficient for (ED).*

In particular, for $d \leq 3$ (ED) is fulfilled if every angle in $\mathcal{T}_h$ is $\leq \pi/2$.

Lemma 2.4.5. *Let $\mathbf{U} \in V_h(\Omega, \mathbb{R}^m)$ be given, and $d \leq 2$. If $S \in V_h(\Omega)$ minimises $E_h(\mathbf{U}, \cdot)$, then $-1 \leq S \leq 1$.*

Proof. For $a \in \mathbb{R}$ define $\overline{a} := -1 \vee a \wedge 1$. Note that for this result it is crucial that we have piecewise affine finite element functions.

Step 1: *If $a, b \in \mathbb{R}$, then $\left(\overline{a} + \overline{b}\right)^2 \leq (a+b)^2$ and $\left(\overline{a} - \overline{b}\right)^2 \leq (a-b)^2$.*
A case differentiation gives

- $a, b \in [-1, 1]$ is trivial.
- $a, b > 1$ or $a, b < -1 \Longrightarrow \left(\overline{a} + \overline{b}\right)^2 = 2^2 \leq (a+b)^2$.
- $a > 1, b < -1 \Longrightarrow \left(\overline{a} + \overline{b}\right)^2 = 0 \leq (a+b)^2$,
  and $b > 1, a < -1$ is symmetrical.
- $a \notin [-1, 1], b \in [-1, 1] \Longrightarrow 0 \leq 1 + \mathrm{sign}(ab)|b| \leq |a| + \mathrm{sign}(ab)|b|$,
  $\Longrightarrow \left(\overline{a} + \overline{b}\right)^2 = (1 + \mathrm{sign}(ab)|b|)^2 \leq (|a| + \mathrm{sign}(ab)|b|)^2 = (a+b)^2$,
  and $b \notin [-1, 1], a \in [-1, 1]$ is symmetrical.

Therefore $\left(\overline{a} + \overline{b}\right)^2 \leq (a+b)^2$, and $\left(\overline{a} - \overline{b}\right)^2 \leq (a-b)^2$ follows by symmetry.

Step 2: *We have $-1 \leq S \leq 1$.*
In case $-1 \leq S \leq 1$ should not be true, we replace $S(\mathbf{x}) = \sum_{\mathbf{z}\in\mathcal{N}} S(\mathbf{z})\varphi_{\mathbf{z}}(\mathbf{x})$ by

$$\overline{S}(\mathbf{x}) := \sum_{\mathbf{z}\in\mathcal{N}} \left(-1 \vee S(\mathbf{z}) \wedge 1\right) \varphi_{\mathbf{z}}(\mathbf{x}) = \mathcal{I}_h(-1 \vee S \wedge 1)\,,$$

for which clearly $-1 \leq \overline{S} \leq 1$. We shall prove $E_h\big(\mathbf{U}, \overline{S}\big) \leq E_h(\mathbf{U}, S)$, by showing energy-decrease for every term involving S, on every triangle $T \in \mathcal{T}_h$. Since $\nabla\mathbf{U}$ is constant on every T, the terms we have to look at are $\int_T S^2 \mathrm{d}\mathbf{x}$, $\int_T |\nabla S|^2 \,\mathrm{d}\mathbf{x}$, and $\int_T \mathcal{I}_h((1-S)^2)\,\mathrm{d}\mathbf{x}$. Let the values of S at the nodal points of T be $S_0, \ldots, S_d$, let $\overline{S}_0, \ldots, \overline{S}_d$ be the corresponding values of $\overline{S}$, let $\varphi_0, \ldots, \varphi_d$ be the corresponding nodal basis functions, and

$\mathbf{x} := (x_1, \ldots, x_d)$. By a simple transformation argument, we can restrict ourselves to the standard simplex, which we shall still call T. Then

$$S(\mathbf{x})|_T = S_0 + \sum_{i=1}^{d} (S_i - S_0)\, x_i,$$

and

$$\nabla S(\mathbf{x})|_T = (S_1 - S_0, \ldots, S_d - S_0)\,,$$

For the first term, a calculation yields

$$\int_T S^2 \mathrm{d}\mathbf{x} = \frac{2}{(d+2)!} \sum_{i=0}^{d} S_i \sum_{j=i}^{d} S_j. \tag{2.4.3}$$

If $d = 1$, then, by Step 1,

$$\begin{aligned}
\int_T \overline{S}^2 \mathrm{d}\mathbf{x} &= \frac{1}{3}\left(\overline{S}_0^2 + \overline{S}_0\overline{S}_1 + \overline{S}_1^2\right) \\
&= \frac{1}{6}\left(\left(\overline{S}_0 + \overline{S}_1\right)^2 + \overline{S}_0^2 + \overline{S}_1^2\right) \\
&\leq \frac{1}{6}\left(\left(S_0 + S_1\right)^2 + S_0^2 + S_1^2\right) \\
&= \int_T S^2 \mathrm{d}\mathbf{x}.
\end{aligned}$$

Similarly, if $d = 2$,

$$\begin{aligned}
\int_T \overline{S}^2 \mathrm{d}\mathbf{x} &= \frac{1}{12}\left(\overline{S}_0^2 + \overline{S}_1^2 + \overline{S}_2^2 + \overline{S}_0\overline{S}_1 + \overline{S}_0\overline{S}_2 + \overline{S}_1\overline{S}_2\right) \\
&= \frac{1}{24}\left(\left(\overline{S}_0 + \overline{S}_1\right)^2 + \left(\overline{S}_0 + \overline{S}_2\right)^2 + \left(\overline{S}_1 + \overline{S}_2\right)^2\right) \\
&\leq \frac{1}{24}\left(\left(S_0 + S_1\right)^2 + \left(S_0 + S_2\right)^2 + \left(S_1 + S_2\right)^2\right) \\
&= \int_T S^2 \mathrm{d}\mathbf{x}.
\end{aligned}$$

Note: Both arguments break down for $d \geq 3$; in fact, counter-examples are easy to find, c.f. Remark 2.4.6.

The second term gives, by Step 1 and symmetry,

$$\begin{aligned}
\int_T |\nabla \overline{S}|^2 \mathrm{d}\mathbf{x} &= \int_T \left(\overline{S}_1 - \overline{S}_0, \ldots, \overline{S}_d - \overline{S}_0\right)^2 \mathrm{d}\mathbf{x} \\
&= \frac{1}{d!}\left(\left(\overline{S}_1 - \overline{S}_0\right)^2 + \cdots + \left(\overline{S}_d - \overline{S}_0\right)^2\right) \\
&\leq \frac{1}{d!}\left(\left(S_1 - S_0\right)^2 + \cdots + \left(S_d - S_0\right)^2\right) \\
&= \int_T |\nabla S|^2.
\end{aligned}$$

As for the last term, again by Step 1,

$$\begin{aligned}\int_T \mathcal{I}_h\big((1-\overline{S})^2\big)\,\mathrm{d}\mathbf{x} &= \sum_{i=1}^{d+1}\big(1-\overline{S}_i\big)^2\int_T \varphi_i \mathrm{d}\mathbf{x}\\ &\le \sum_{i=1}^{d+1}(1-S_i)^2\int_T\varphi_i\mathrm{d}\mathbf{x} = \int_T \mathcal{I}_h\big((1-S)^2\big)\,\mathrm{d}\mathbf{x}.\end{aligned}$$

□

Remark 2.4.6. *For $d=3$, Step 2 in the above proof is wrong: Let $S_0 := S_1 := S_2 := 1$, and $S_3 := -3/2$. Then, by* (2.4.3)*,*

$$\int_T \overline{S}^2\mathrm{d}\mathbf{x} = \frac{1}{60}\sum_{i=0}^{d}\overline{S}_i\sum_{j=i}^{d}\overline{S}_j = \frac{1}{15},$$

while

$$\int_T S^2\mathrm{d}\mathbf{x} = \frac{1}{60}\sum_{i=0}^{d}S_i\sum_{j=i}^{d}S_j = \frac{1}{16}.$$

We suspect that there exist dimension-dependent constants c_d, at which one could crop $|S|$, so that the energy is still decreasing (also replacing $(1-s)^2$ by $(c_d-s)^2$).

Lemma 2.4.7. *Let $\mathcal{T}_h$ be a quasi-uniform triangulation of Ω satisfying (ED), $\varrho>0$ fixed, $S_0\in V_h(\Omega)$, and $\mathbf{U}_0\in H^1_h(\mathcal{T}_h)$. Then Algorithm 2.4.1 terminates within a finite number of iterations with output $(\mathbf{U},S)\in H^1_h(\mathcal{T}_h)\times V_h(\Omega,[-1,1])$ and $\mathbf{W}\in V_h(\Omega,\mathbb{R}^m)$ such that $\|\nabla\mathbf{W}\|\le\varrho$, and $E_h(\mathbf{U},S)\le E_h(\mathbf{U}_0,S_0)$.*

Proof. We proceed by induction. Suppose that for some $n\ge 0$ we have $(\mathbf{U}_n,S_n)\in H^1_h(\mathcal{T}_h)\times V_h(\Omega)$. The set K^n_h is a subspace of $V_h(\Omega,\mathbb{R}^m)$. Therefore, by Lax-Milgram, there is a unique $\mathbf{W}_n\in K^n_h$ such that (2.4.1) is fulfilled. Since $\mathbf{W}_n(\mathbf{z})\cdot\mathbf{U}_n(\mathbf{z})=0$ and $|\mathbf{U}_n(\mathbf{z})|=1$, we have $|\mathbf{U}_n(\mathbf{z})-\mathbf{W}_n(\mathbf{z})|\ge 1$ for $\mathbf{z}\in\mathcal{N}$; i.e., $\mathbf{U}_{n+1}$ is well-defined and in $H^1_h(\mathcal{T}_h)$. And since $\mathbf{0}\in K^n_h$ and $\mathcal{T}_h$ fulfils (ED), we get

$$E_h(\mathbf{U}_{n+1},S_n)\le E_h(\mathbf{U}_n-\mathbf{W}_n,S_n)\,.$$

Step 4 of Algorithm 2.4.1 has a solution S_{n+1} by convexity and coercivity of the functional. So

$$E_h(\mathbf{U}_{n+1},S_{n+1})\le E_h(\mathbf{U}_{n+1},S_n)\le E_h(\mathbf{U}_n-\mathbf{W}_n,S_n)\le E_h(\mathbf{U}_n,S_n)\,.$$

In fact, $E_h(\mathbf{U}_{n+1},S_{n+1})\le E_h(\mathbf{U}_{n+1},W)$ for all $W\in V_h(\Omega)$. Therefore, by Lemma 2.4.5, we can assume $-1\le S_{n+1}\le 1$.

Recycling the proof of Lemma 2.3.9 we have

$$0\le\frac{1}{2}\int_\Omega\gamma\left(S_n^2+k_\varepsilon\right)|\nabla\mathbf{W}_n|^2+\lambda|\mathbf{W}_n|^2\,\mathrm{d}\mathbf{x}\le\widetilde{E}(\mathbf{U}_n,S_n)-\widetilde{E}(\mathbf{U}_{n+1},S_{n+1})\,.$$

This implies, just like in the continuous case (see proof of Proposition 2.3.10), that $\|\mathbf{W}_n\|_{H^1(\Omega;\mathbb{R}^m)}\le\varrho$ for n large enough. □

Theorem 2.4.8. *Let $\{\mathcal{T}_{h_l}\}$ be a sequence of quasi-uniform triangulations satisfying (ED) with maximal mesh size $h_l \to 0$ for $l \to +\infty$, $\varrho_l \to 0$ for $l \to +\infty$, and $E_{h_l}(\mathbf{U}_0, S_0) \leq C_0 < +\infty$ independently of h_l. Let $\{\mathbf{U}_l, S_l\}$ be the output of Algorithm 2.4.1 (after termination) from input $(\mathbf{U}_l^0, S_l^0, \varrho_l)$. Then the sequence $\{\mathbf{U}_l, S_l\}$ converges weakly in $H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega)$ (up to subsequences, not relabelled) for $l \to +\infty$ to a point $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega, [-1, 1])$, with $AT_\varepsilon(\mathbf{u}, s) \leq \liminf_l AT_\varepsilon(\mathbf{U}_l, S_l) \leq \liminf_l AT_\varepsilon(\mathbf{U}_l^0, S_l^0)$.*

Proof. By assumption and Lemma 2.4.7, we have

$$E_{h_l}(\mathbf{U}_l, S_l) \leq E_{h_l}\big(\mathbf{U}_l^0, S_l^0\big) \leq C_0,$$

and $-1 \leq S_l \leq 1$. This implies, like in the proof of Proposition 2.3.10, uniform boundedness of H^1-norms of iterates $\mathbf{U}_l$ and S_l. Hence we can extract a subsequence that converges weakly in $H^1 \times H^1$ to some map $(\mathbf{u}, s)$. Poincaré's inequality (elementwise), $|\mathbf{U}_l(\mathbf{z})| = 1$ for all $\mathbf{z} \in \mathcal{N}_{h_l}$, and $|\mathbf{U}_l| \leq 1$ a.e. imply

$$\big\|\,|\mathbf{U}_l|^2 - 1\big\| \leq C h_l \left\|2\mathbf{U}_l^T \nabla \mathbf{U}_l\right\| \leq C h_l.$$

So $\mathbf{U}_l \to \mathbf{u}$ a.e. leads to $|\mathbf{u}| = 1$ a.e.

Since $H^1(\Omega)$ is a Hilbert space and $\{\varphi \in H^1(\Omega) : 0 \leq \varphi \leq 1 \text{ a.e}\} \subset H^1(\Omega)$ is a closed, convex set, it is weakly closed. Therefore, by the weak convergence in H^1 of $S_l \rightharpoonup s$, we get $-1 \leq s \leq 1$.

Finally, by weak lower semicontinuity of $AT_\varepsilon(\cdot, \cdot)$,

$$\begin{aligned}
AT_\varepsilon(\mathbf{u}, s) &\leq \liminf_l AT_\varepsilon(\mathbf{U}_l, S_l) \\
&\leq \liminf_l \Big(E_{h_l}(\mathbf{U}_l, S_l) + c \left\|\mathcal{I}_h\big((1 - S_l)^2\big) - (1 - S_l)^2\right\|_{L^1(\Omega)} \Big) \\
&\leq \liminf_l \Big(E_{h_l}\big(\mathbf{U}_l^0, S_l^0\big) + c h_l \left\|S_l\right\|_{L^2(\Omega)} \left\|\nabla S_l\right\|_{L^2(\Omega)} \Big) \\
&\leq \liminf_l E_{h_l}\big(\mathbf{U}_l^0, S_l^0\big) \\
&\leq \liminf_l \Big(AT_\varepsilon\big(\mathbf{U}_l^0, S_l^0\big) + c h_l \left\|S_l^0\right\|_{L^2(\Omega)} \left\|\nabla S_l^0\right\|_{L^2(\Omega)} \Big) \\
&\leq \liminf_l AT_\varepsilon\big(\mathbf{U}_l^0, S_l^0\big)\,.
\end{aligned}$$

□

Remark 2.4.9. *We have essentially the same problem as in the continuous case (c.f. Remark 2.3.11); i.e., we cannot show that $(\mathbf{u}, s)$ is a stationary point.*

2.5 Γ-Convergence for Penalisation & Splitting

In order to resolve the problems with passing to the limit, we now use a penalisation approach instead of projection. This requires adding a term to the Ambrosio-Tortorelli energy, which penalises the sphere constraint. In this section, we show that this addition does not affect Γ-convergence to the Mumford-Shah functional, if the penalisation term is properly scaled.

Let $\Omega \subset \mathbb{R}^d$, γ, α, λ be fixed positive constants, $\varepsilon, \delta_\varepsilon > 0$, $k_\varepsilon \geq 0$, $\mathbf{g} \in L^\infty(\Omega, \mathbb{S}^{m-1})$, and $G_\varepsilon, G : L^2(\Omega, \mathbb{R}^m) \times L^2(\Omega) \to [0, +\infty]$ be defined by

$$G_\varepsilon(\mathbf{u}, s) := \begin{cases} \dfrac{\gamma}{2}\displaystyle\int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \dfrac{\lambda}{2}\displaystyle\int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x} & \text{if } \mathbf{u} \in H^1(\Omega, \mathbb{R}^m)\,, \\ \quad +\alpha \displaystyle\int_\Omega \left(\varepsilon |\nabla s|^2 + \dfrac{(1-s)^2}{4\varepsilon}\right) \mathrm{d}\mathbf{x} & s \in H^1(\Omega, [0,1]), \\ \quad +\dfrac{1}{4\delta_\varepsilon}\displaystyle\int_\Omega \left(|\mathbf{u}|^2 - 1\right)^2 \mathrm{d}\mathbf{x} & \\ +\infty, & \text{otherwise,} \end{cases}$$

and

$$G(\mathbf{u}, s) := \begin{cases} \dfrac{\gamma}{2}\displaystyle\int_\Omega |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \alpha \mathcal{H}^{d-1}(S_{\mathbf{u}}) + \dfrac{\lambda}{2}\displaystyle\int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x} & \text{if } \mathbf{u} \in GSBV(\Omega, \mathbb{S}^{m-1}) \\ & \text{and } s = 1 \text{ a.e.} \\ +\infty, & \text{otherwise.} \end{cases}$$

Theorem 2.5.1. *If $\Omega \subset \mathbb{R}^d$ is open and bounded with Lipschitz boundary, $\delta_\varepsilon \underset{\varepsilon \to 0}{\longrightarrow} 0$, $k_\varepsilon = o(\varepsilon)$, and $k_\varepsilon = o(\delta_\varepsilon)$, then $G_\varepsilon \xrightarrow[\varepsilon \to 0]{\Gamma} G$ in $L^2(\Omega, \mathbb{R}^m) \times L^2(\Omega)$.*

Moreover, there exists a solution $\{\mathbf{u}_\varepsilon, s_\varepsilon\}$ to the minimum problem

$$m_\varepsilon = \inf_{\substack{\mathbf{u} \in H^1(\Omega, \mathbb{R}^m), \\ s \in H^1(\Omega, [0,1])}} G_\varepsilon(\mathbf{u}, s)$$

with $\|\mathbf{u}_\varepsilon\|_{L^\infty} \leq C$, and every cluster point of $\{\mathbf{u}_\varepsilon\}$ is a solution to the minimum problem

$$m = \inf_{\mathbf{u} \in GSBV(\Omega, \mathbb{S}^{m-1})} G(\mathbf{u}, 1),$$

and $m_\varepsilon \to m$ as $\varepsilon \to 0^+$.

For the lim inf inequality we can apply the work of Focardi ([62, Lemma 3.3]). For the lim sup inequality we use the same construction as Ambrosio and Tortorelli in [4], so it is enough to verify that the penalisation term we added vanishes for $\varepsilon \to 0^+$. This is explained in more detail below.

Proof. For notational convenience, we first localise the functionals above, denoting by $G_\varepsilon(\mathbf{u}, s, A)$ and $G(\mathbf{u}, s, A)$ the same functionals with integration over $A \subseteq \Omega$ instead of Ω, and $\mathcal{H}^{d-1}(S_{\mathbf{u}})$ replaced by $\mathcal{H}^{d-1}(S_{\mathbf{u}} \cap A)$.

Step 1: *The Liminf Inequality.*

Let $\varepsilon \to 0^+$, and $(\mathbf{u}_\varepsilon, s_\varepsilon) \to (\mathbf{u}, s)$ in $L^2(\Omega, \mathbb{R}^m) \times L^2(\Omega)$. Up to subsequences, we can suppose that $(\mathbf{u}_\varepsilon, s_\varepsilon) \to (\mathbf{u}, s)$ a.e., and that $\lim_{\varepsilon \to 0^+} G_\varepsilon(\mathbf{u}_\varepsilon, s_\varepsilon)$ exists and is finite. We can further assume $s = 1$ a.e., since otherwise $\int_\Omega (1 - s_\varepsilon)^2 \, \mathrm{d}\mathbf{x} \nrightarrow 0$, and $G_\varepsilon(\mathbf{u}_\varepsilon, s_\varepsilon) \to \infty$. Similarly, we get $|\mathbf{u}|^2 = 1$ a.e.

We now have to show

$$\liminf_{\varepsilon \to 0^+} G_\varepsilon(\mathbf{u}_\varepsilon, s_\varepsilon) \geq G(\mathbf{u}, s).$$

Since it is clear that $\int_\Omega |\mathbf{u}_\varepsilon - \mathbf{g}|^2 \, \mathrm{d}\mathbf{x} \to \int_\Omega |\mathbf{u} - \mathbf{g}|^2 \, \mathrm{d}\mathbf{x}$, and that the penalisation term is non-negative, it is sufficient to prove that $\mathbf{u} \in GSBV(\Omega, \mathbb{R}^m)$, and

$$\begin{aligned} &\liminf_{\varepsilon \to 0^+} \int_\Omega \left(s_\varepsilon^2 + k_\varepsilon\right) |\nabla \mathbf{u}_\varepsilon|^2 \, \mathrm{d}\mathbf{x} + 2 \int_\Omega \left(\varepsilon |\nabla s_\varepsilon|^2 + \frac{(1 - s_\varepsilon)^2}{4\varepsilon} \right) \mathrm{d}\mathbf{x} \\ &\geq \int_\Omega |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + 2\mathcal{H}^{d-1}(S_\mathbf{u}) \, . \end{aligned}$$

This was shown for a more general situation in [62, Lemma 3.3] (see also [63]).

Step 2: *The Limsup Inequality.*

It suffices to consider the case $\mathbf{u} \in SBV(\Omega, \mathbb{R}^m) \cap L^\infty(\Omega, \mathbb{R}^m)$. We can also assume $\nabla \mathbf{u} \in L^2\big(\Omega, \mathbb{R}^{d \times m}\big)$, $|\mathbf{u}|^2 = 1$ a.e., and (see [63, Theorem 2.7.14]) that $S_\mathbf{u}$ is essentially closed in Ω; i.e., $\mathcal{H}^{d-1}\big(\Omega \cap \big(\overline{S_\mathbf{u}} \setminus S_\mathbf{u}\big)\big) = 0$. Setting $d(\mathbf{x}) := \mathrm{dist}(\mathbf{x}, S_\mathbf{u})$, we define the *Minkowski content of* $S_\mathbf{u}$

$$\mathcal{M}^{d-1}(S_\mathbf{u}) := \lim_{\delta \to 0^+} \mathcal{M}_\delta^{d-1}(S_\mathbf{u}) := \lim_{\delta \to 0^+} \frac{|\{\mathbf{x} \in \Omega : d(\mathbf{x}) < \delta\}|}{2\delta}.$$

It is well-known that for $S_\mathbf{u}$ essentially closed,

$$\lim_{\delta \to 0^+} \mathcal{M}_\delta^{d-1}(S_\mathbf{u}) = \mathcal{H}^{d-1}(S_\mathbf{u}) \tag{2.5.1}$$

(see [61, Theorem 3.2.39]). So, there exists a sequence $w_\varepsilon \to 0^+$, such that

$$|\{\mathbf{x} \in \Omega : d(\mathbf{x}) < \delta\}| \leq 2\delta \left(\mathcal{H}^{d-1}(S_\mathbf{u}) + w_\varepsilon\right), \tag{2.5.2}$$

for every $\delta \geq 0$ small enough.

Given such functions $\mathbf{u}$, and $s = 1$ a.e., we have to construct $\{\mathbf{u}_\varepsilon, s_\varepsilon\}$ that converge in $L^2(\Omega, \mathbb{R}^m) \times L^2(\Omega)$ to $(\mathbf{u}, s)$, such that

$$\limsup_{\varepsilon \to 0^+} G_\varepsilon(\mathbf{u}_\varepsilon, s_\varepsilon) \leq G(u, s)$$

for any positive sequence ε converging to zero.

It is natural to require $s_\varepsilon \equiv 0$ in some ε-dependent neighbourhood of $S_\mathbf{u}$, s_ε converging to 1 everywhere outside a larger neighbourhood of $S_\mathbf{u}$, and smooth in between, as well as $\mathbf{u}_\varepsilon \equiv \mathbf{u}$ everywhere outside some neighbourhood of $S_\mathbf{u}$.

To this end, we use the same construction as in the paper [4] by Ambrosio and Tortorelli: Choose a positive sequence b_ε, such that $b_\varepsilon = o(\varepsilon)$, $b_\varepsilon = o(\delta_\varepsilon)$, and $k_\varepsilon = o(b_\varepsilon)$.

For any $b > 0$, set $S_b := \{\mathbf{x} \in \Omega : d(\mathbf{x}) < b\}$. Thanks to (2.5.2), $|S_b| = O(b)$. For $t \geq b_\varepsilon$, let

$$\begin{aligned} \sigma_\varepsilon(t) &:= 1 - \exp\left(-\frac{t - b_\varepsilon}{2\varepsilon}\right), \text{ so that} \\ \sigma'_\varepsilon(t) &= \frac{1}{2\varepsilon}\exp\left(-\frac{t - b_\varepsilon}{2\varepsilon}\right). \end{aligned}$$

We now set (c.f. Figure 2.1)

$$s_\varepsilon(\mathbf{x}) := \begin{cases} 0 & \text{if } \mathbf{x} \in S_{b_\varepsilon}, \\ \sigma_\varepsilon(d(\mathbf{x})) & \text{if } \mathbf{x} \in S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \setminus S_{b_\varepsilon}, \\ 1 - \varepsilon & \text{if } \mathbf{x} \in \Omega \setminus S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}}, \end{cases} \tag{2.5.3}$$

and

$$\mathbf{u}_\varepsilon(\mathbf{x}) := \mathbf{u}(\mathbf{x}) \min\left\{\frac{d(\mathbf{x})}{b_\varepsilon}, 1\right\}.$$

Note that $0 < 2\varepsilon \ln \frac{1}{\varepsilon} \to 0^+$, and $\varepsilon = o\big(2\varepsilon \ln \frac{1}{\varepsilon}\big)$.

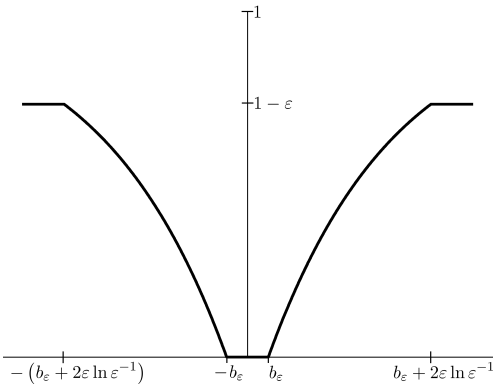

Figure 2.1: Sketch of $s_\varepsilon(\mathbf{x})$ in the case $S_\mathbf{u} = \{0\}$, and $d = 1$.

By construction, $(\mathbf{u}_\varepsilon, s_\varepsilon) \to (\mathbf{u}, 1)$ in $L^2(\Omega, \mathbb{R}^m) \times L^2(\Omega)$, as $\varepsilon \to 0^+$.

Therefore, we have for the term penalising the sphere constraint,

$$\frac{1}{4\delta_\varepsilon} \int_\Omega \big(|\mathbf{u}_\varepsilon|^2 - 1\big)^2 \, \mathrm{d}\mathbf{x} \leq c \frac{|S_{b_\varepsilon}|}{\delta_\varepsilon} \leq c \frac{b_\varepsilon}{\delta_\varepsilon} \to 0. \tag{2.5.4}$$

So this term does not contribute to the lim sup. This calculation motivates why we cannot expect good experimental results for δ_ε too small (compared to b_ε, which in turn is between ε and k_ε); i.e., we have to sacrifice something in terms of the sphere constraint, c.f. our experiments in Section 2.8.4.

The other terms are just like in the original paper [4]. We repeat the corresponding calculations for the convenience of the reader.

By construction,

$$\nabla \mathbf{u}_\varepsilon(\mathbf{x}) = \begin{cases} d(\mathbf{x})/b_\varepsilon \nabla \mathbf{u}(\mathbf{x}) + \mathbf{u}(\mathbf{x}) \otimes \nabla d(\mathbf{x})/b_\varepsilon & \text{if } \mathbf{x} \in S_{b_\varepsilon}, \\ \nabla \mathbf{u}(\mathbf{x}) & \text{otherwise.} \end{cases}$$

And since $|\nabla d| = 1$ a.e. (see e.g. [61, Proof of Lemma 3.2.34]), $|\nabla \mathbf{u}_\varepsilon(\mathbf{x})| \leq |\nabla \mathbf{u}(\mathbf{x})| + \|\mathbf{u}\|_{L^\infty}/b_\varepsilon$. So we get

$$\begin{aligned}\int_\Omega \left(s_\varepsilon^2 + k_\varepsilon\right) |\nabla \mathbf{u}_\varepsilon|^2 \,\mathrm{d}\mathbf{x} &\leq \left[\int_{\Omega\setminus S_{b_\varepsilon}} + \int_{S_{b_\varepsilon}}\right] \left(s_\varepsilon^2 + k_\varepsilon\right) |\nabla \mathbf{u}_\varepsilon|^2 \,\mathrm{d}\mathbf{x} \\ &\leq (1+k_\varepsilon) \int_{\Omega\setminus S_{b_\varepsilon}} |\nabla \mathbf{u}|^2 \,\mathrm{d}\mathbf{x} \\ &\quad + Ck_\varepsilon \left(\int_{S_{b_\varepsilon}} |\nabla \mathbf{u}|^2 \,\mathrm{d}\mathbf{x} + \frac{|S_{b_\varepsilon}|\,\|\mathbf{u}\|_{L^\infty}^2}{b_\varepsilon^2}\right).\end{aligned}$$

Since $|S_{b_\varepsilon}| = O(b_\varepsilon)$, and $k_\varepsilon = o(b_\varepsilon)$,

$$\limsup_{\varepsilon\to 0^+} \frac{\gamma}{2} \int_\Omega \left(s_\varepsilon^2 + k_\varepsilon\right) |\nabla \mathbf{u}_\varepsilon|^2 \,\mathrm{d}\mathbf{x} \leq \frac{\gamma}{2} \int_\Omega |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x}. \tag{2.5.5}$$

The next term gives

$$\begin{aligned}&\int_\Omega \left(\varepsilon |\nabla s_\varepsilon|^2 + \frac{(1-s_\varepsilon)^2}{4\varepsilon}\right) \mathrm{d}\mathbf{x} \\ &\quad = \left[\int_{S_{b_\varepsilon}} + \int_{S_{b_\varepsilon + 2\varepsilon\ln\frac{1}{\varepsilon}}\setminus S_{b_\varepsilon}} + \int_{\Omega\setminus S_{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}}}\right] \left(\varepsilon\,|\nabla s_\varepsilon|^2 + \frac{(1-s_\varepsilon)^2}{4\varepsilon}\right) \mathrm{d}\mathbf{x} \\ &\quad =: I_\varepsilon + II_\varepsilon + III_\varepsilon.\end{aligned}$$

We shall show that I_ε and III_ε vanish. But first, set $g(t) := |\{\mathbf{x}\in\Omega : d(\mathbf{x}) < t\}|$, which is a primitive of $\mathcal{H}^{d-1}(\{\mathbf{x}\in\Omega : d(\mathbf{x}) = t\})$. Using the coarea formula (see e.g. [3, Theorem 2.93]) and $|\nabla d| = 1$ a.e., we get

$$\begin{aligned}II_\varepsilon &= \int_{b_\varepsilon}^{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}} \int_{\{\mathbf{y}\in\Omega : d(\mathbf{y})=t\}} \left(\varepsilon\,|\nabla s_\varepsilon(\mathbf{y})|^2 + \frac{(1-s_\varepsilon(\mathbf{y}))^2}{4\varepsilon}\right) |\nabla d(\mathbf{y})|^{-1} \,\mathrm{d}\mathcal{H}^{d-1}(\mathbf{y})\,\mathrm{d}t \\ &\leq \int_{b_\varepsilon}^{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}} \left(\varepsilon\,(\sigma_\varepsilon'(t))^2 + \frac{(1-\sigma_\varepsilon(t))^2}{4\varepsilon}\right) \mathcal{H}^{d-1}(\{\mathbf{y}\in\Omega : d(\mathbf{y}) = t\})\,\mathrm{d}t \\ &= \int_{b_\varepsilon}^{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}} \left(\varepsilon\,(\sigma_\varepsilon'(t))^2 + \frac{(1-\sigma_\varepsilon(t))^2}{4\varepsilon}\right) g'(t)\mathrm{d}t \\ &= \frac{1}{2\varepsilon} \int_{b_\varepsilon}^{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}} \exp\left(-\frac{t-b_\varepsilon}{\varepsilon}\right) g'(t)\mathrm{d}t.\end{aligned}$$

Using integration by parts, (2.5.2), and integration by parts again,

$$\begin{aligned} II_\varepsilon &\leq \frac{1}{2\varepsilon}\left(\varepsilon^2 g\left(b_\varepsilon + 2\varepsilon\ln\frac{1}{\varepsilon}\right) - g(b_\varepsilon) + \frac{1}{\varepsilon}\int_{b_\varepsilon}^{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}} \exp\left(-\frac{t-b_\varepsilon}{\varepsilon}\right) g(t)\mathrm{d}t\right) \\ &\leq \frac{\mathcal{H}^{d-1}(S_{\mathbf{u}}) + w_\varepsilon}{\varepsilon}\left(\varepsilon^2\left(b_\varepsilon + 2\varepsilon\ln\frac{1}{\varepsilon}\right) + \frac{1}{\varepsilon}\int_{b_\varepsilon}^{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}} \exp\left(-\frac{t-b_\varepsilon}{\varepsilon}\right) t\mathrm{d}t\right) \\ &\leq \frac{C}{\varepsilon}\left(\varepsilon^2\left(b_\varepsilon + 2\varepsilon\ln\frac{1}{\varepsilon}\right) - \varepsilon^2\left(b_\varepsilon + 2\varepsilon\ln\frac{1}{\varepsilon}\right) + b_\varepsilon\right) \\ &\quad + \frac{\mathcal{H}^{d-1}(S_{\mathbf{u}}) + w_\varepsilon}{\varepsilon}\int_{b_\varepsilon}^{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}} \exp\left(-\frac{t-b_\varepsilon}{\varepsilon}\right)\mathrm{d}t \\ &= C\frac{b_\varepsilon}{\varepsilon} + \left(\mathcal{H}^{d-1}(S_{\mathbf{u}}) + w_\varepsilon\right)\left(1-\varepsilon^2\right), \end{aligned}$$

so,

$$\limsup_{\varepsilon\to 0^+} II_\varepsilon \leq \mathcal{H}^{d-1}(S_{\mathbf{u}}),$$

and what is left to show is $I_\varepsilon, III_\varepsilon \to 0$. Indeed, by the construction of s_ε, $|S_{b_\varepsilon}| = O(b_\varepsilon)$, and $b_\varepsilon = o(\varepsilon)$,

$$I_\varepsilon \leq \int_{S_{b_\varepsilon}} \frac{1}{4\varepsilon}\mathrm{d}\mathbf{x} \leq C\frac{b_\varepsilon}{\varepsilon} \to 0.$$

And similarly,

$$III_\varepsilon = \int_{\Omega\setminus S_{b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}}} \frac{\varepsilon^2}{4\varepsilon}\mathrm{d}\mathbf{x} \to 0,$$

so, indeed,

$$\limsup_{\varepsilon\to 0^+} \int_\Omega \left(\varepsilon\left|\nabla s_\varepsilon\right|^2 + \frac{\left(1-s_\varepsilon\right)^2}{4\varepsilon}\right)\mathrm{d}\mathbf{x} \leq \mathcal{H}^{d-1}(S_{\mathbf{u}}), \tag{2.5.6}$$

Inequalities (2.5.5), (2.5.6), and (2.5.4) together prove the lim sup inequality.

Step 3: *Convergence of Minimisers.*

The functional G_ε is coercive and lower semicontinuous in L^2. So for every $\varepsilon > 0$ there exists a minimising pair $(\mathbf{u}_\varepsilon, s_\varepsilon)$ of G_ε. By a simple truncation argument, $\|\mathbf{u}_\varepsilon\|_{L^\infty} \leq C$. As above, we can assume that $(\mathbf{u}_\varepsilon, s_\varepsilon) \in SBV(\Omega,\mathbb{R}^m)\times SBV(\Omega)\cap L^\infty(\Omega,\mathbb{R}^m)\times L^\infty(\Omega)$. By the SBV Closure and Compactness Theorems [3, Theorems 4.7 and 4.8], there exists a subsequence $\left\{\mathbf{u}_{\varepsilon_j}, s_{\varepsilon_j}\right\}$ converging to some $(\mathbf{u},1)$ in $L^2(\Omega,\mathbb{R}^m)\times L^2(\Omega)$, with $\mathbf{u}\in SBV(\Omega,\mathbb{R}^m)$. Thus, the stability of minimising sequences under Γ-convergence (Lemma 2.2.1(3)) concludes the proof. □

2.6 Algorithm for Penalisation & Splitting

2.6.1 Algorithm

Let $\Omega \subset \mathbb{R}^d$, be a polyhedral Lipschitz domain, and let $\mathbf{g} : \Omega \to \mathbb{S}^{m-1}$ be the chromaticity component of a given image. For $\mathbf{u}, \mathbf{g} \in H^1(\Omega, \mathbb{R}^m)$, $s \in H^1(\Omega, [0,1])$, and $0 < \varepsilon, k_\varepsilon, \delta_\varepsilon \ll 1$, we want to minimise the following vector valued Ambrosio-Tortorelli-Ginzburg-Landau energy using a splitting strategy:

$$\begin{aligned} G_\varepsilon(\mathbf{u}, s) &= \frac{\gamma}{2}\int_\Omega \left(s^2 + k_\varepsilon\right)|\nabla \mathbf{u}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega |\mathbf{u} - \mathbf{g}|^2 \,\mathrm{d}\mathbf{x} \\ &\quad + \alpha \int_\Omega \varepsilon |\nabla s|^2 + \frac{1}{4\varepsilon}(1-s)^2 \mathrm{d}\mathbf{x} + \frac{1}{4\delta_\varepsilon}\int_\Omega \left(|\mathbf{u}|^2 - 1\right)^2 \mathrm{d}\mathbf{x}. \end{aligned} \tag{2.6.1}$$

In this section, we shall always assume $\gamma, \alpha, \varepsilon, k_\varepsilon, \delta_\varepsilon$ to be fixed and positive, $\lambda \geq 0$, and $d \leq 2$ (the last assumption is again only used to show that iterates $S_n \in [-1,1]$, and that their weak limit $s \in [0,1]$).

Definition 2.6.1. *A tuple* $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega, [0,1])$ *is called a* weak solution *to the problem* $\inf G_\varepsilon$*, if and only if*

$$\gamma\left(\left(s^2 + k_\varepsilon\right)\nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\right) + \lambda\left(\mathbf{u} - \mathbf{g}, \boldsymbol{\varphi}\right) + \frac{1}{\delta_\varepsilon}\left(\left(|\mathbf{u}|^2 - 1\right)\mathbf{u}, \boldsymbol{\varphi}\right) = 0 \tag{2.6.2}$$

for all $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$*, and*

$$2\alpha\varepsilon\left(\nabla s, \nabla \varphi\right) + \gamma\left(|\nabla \mathbf{u}|^2 s, \varphi\right) + \frac{\alpha}{2\varepsilon}\left(s - 1, \varphi\right) = 0 \tag{2.6.3}$$

for all $\varphi \in H^1(\Omega) \cap L^\infty(\Omega)$*.*

We use the same finite element setting as in Section 2.4, in particular, we shall always assume the triangulation $\mathcal{T}_h$ to be quasi-uniform. For $\mathbf{U}, \mathbf{G} \in V_h(\Omega, \mathbb{R}^m)$ and $S \in V_h(\Omega, [-1,1])$, let

$$\begin{aligned} G_{\varepsilon,h}(\mathbf{U}, S) &= \frac{\gamma}{2}\int_\Omega \left(S^2 + k_\varepsilon\right)|\nabla \mathbf{U}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega |\mathbf{U} - \mathbf{G}|^2 \,\mathrm{d}\mathbf{x} \\ &\quad + \alpha \int_\Omega \varepsilon |\nabla S|^2 + \frac{1}{4\varepsilon}\mathcal{I}_h\big((1-S)^2\big) \,\mathrm{d}\mathbf{x} + \frac{1}{4\delta_\varepsilon}\int_\Omega \left(|\mathbf{U}|^2 - 1\right)^2 \mathrm{d}\mathbf{x}. \end{aligned} \tag{2.6.4}$$

In the algorithm below we use $\mathbf{G} := \mathbf{r}_h(\mathbf{g}) \in V_h(\Omega, \mathbb{R}^m)$, i.e., the Clément interpolation of $\mathbf{g}$. This allows the use of non-smooth $\mathbf{g}$. If $\mathbf{g} \in C^0(\Omega, \mathbb{R}^m)$, the Lagrange interpolation would do as well.

Algorithm 2.6.2. *Let* $\mathbf{U}_0, \mathbf{G} \in V_h(\Omega, \mathbb{R}^m)$ *and* $S_0 \in V_h(\Omega)$ *be given. For* $n = 1, \ldots$ *until convergence do*

(1) *Compute* $\mathbf{U}_n \in V_h(\Omega, \mathbb{R}^m)$ *such that*

$$\gamma\left(\left(S_{n-1}^2 + k_\varepsilon\right)\nabla \mathbf{U}_n, \nabla \mathbf{W}\right) + \lambda\left(\mathbf{U}_n - \mathbf{G}, \mathbf{W}\right) + \frac{1}{\delta_\varepsilon}\left(\left(|\mathbf{U}_n|^2 - 1\right)\mathbf{U}_n, \mathbf{W}\right) = 0 \tag{2.6.5}$$

for all $\mathbf{W} \in V_h(\Omega, \mathbb{R}^m)$*.*

(2) *Compute* $S_n \in V_h(\Omega)$ *such that*

$$2\alpha\varepsilon\left(\nabla S_n, \nabla W\right) + \gamma\left(S_n\left|\nabla \mathbf{U}_n\right|^2, W\right) + \frac{\alpha}{2\varepsilon}\left(S_n - 1, W\right)_h = 0 \tag{2.6.6}$$

for all $W \in V_h(\Omega)$.

2.6.2 Analysis

We start with a discussion of relevant stability properties of iterates from Algorithm 2.6.2.

Lemma 2.6.3. *Algorithm 2.6.2 decreases* $G_{\varepsilon,h}$ *with respect to* $n \in \mathbb{N}$.

Proof. For any $n \in \mathbb{N}$ fixed, Algorithm 2.6.2 ensures, that

$$G_{\varepsilon,h}(\mathbf{U}_{n+1}, S_{n+1}) \leq G_{\varepsilon,h}(\mathbf{U}_{n+1}, S_n) \leq G_{\varepsilon,h}(\mathbf{U}_n, S_n)\,.$$

□

The following existence and uniqueness result follows by standard coercivity and convexity arguments for $G_{\varepsilon,h}$ (see e.g. [64, Section 8.4]). The fact $-1 \leq S \leq 1$ follows from Lemma 2.4.5.

Proposition 2.6.4. *There exists a function* $\mathbf{U} \in V_h(\Omega, \mathbb{R}^m)$*, such that equation* (2.6.5) *holds for all* $\mathbf{W} \in V_h(\Omega, \mathbb{R}^m)$*, and a unique function* $S \in V_h(\Omega, [-1, 1])$*, such that equation* (2.6.6) *holds for all* $W \in V_h(\Omega)$.

Main convergence properties of iterates from Algorithm 2.6.2 are given in the following

Theorem 2.6.5. *Let* $\{\mathcal{T}_{h_l}\}$ *be a sequence of quasi-uniform triangulations with maximal mesh size* $h_l \to 0$ *for* $l \to +\infty$*, and* $G_{\varepsilon,h_l}\left(\mathbf{U}_0^l, S_0^l\right) \leq C_0 < +\infty$ *independently of* h_l*. Then the sequences* $\left\{\mathbf{U}_m^l, S_m^l\right\}_{m,l}$*, constructed by Algorithm 2.6.2 from inputs* $\left(\mathbf{U}_0^l, S_0^l\right)$ *have a (diagonal) subsequence called* $\{\mathbf{U}_n, S_n\}_n$*, such that* $\mathbf{U}_n$ *converges strongly in* $H^1(\Omega, \mathbb{R}^m)$*, and* S_n *converges weakly in* $H^1(\Omega)$ *to some* $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega, [0, 1])$*, which is a weak solution as in Definition 2.6.1.*

For identifying limits in the proof of Theorem 2.6.5, it will be crucial to prove strong L^2 convergence of $\nabla \mathbf{U}_n$ to $\nabla \mathbf{u}$, for which we use a strategy derived from [25, Proof of Theorem 2], where the authors show convergence of two adaptive, stationary finite element approximations for the minimisation of the unconstrained Ambrosio-Tortorelli energy: In Step 2 we show that $\mathbf{u}$ fulfils equation (2.6.2), then we use equations (2.6.2) and (2.6.5) and dominated convergence (c.f. Lemma 2.6.6, also derived from [25]) to show strong L^2 convergence of $\nabla \mathbf{U}_n$ to $\nabla \mathbf{u}$ in Step 3, and finally we use this to show that s fulfils equation (2.6.3) in Step 4.

Lemma 2.6.6. *Let* $p_n, p \in H^1(\Omega) \cap L^\infty(\Omega)$*, such that* $\|p_n\|_{L^\infty(\Omega)}, \|p\|_{L^\infty(\Omega)} \leq C < +\infty$ *a.e., independently of* n*, and* $p_n \to p$ *in* $L^2(\Omega)$*. Then*

$$\lim_n \left(|p_n - p|, |\nabla \boldsymbol{\varphi}|^2\right) = 0 \qquad \forall\, \boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)\,.$$

Proof. Let $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$ be fixed. Choose a subsequence $\{p_{n_j}\}_j$ of $\{p_n\}_n$ such that

$$\begin{aligned} \lim_j \left(\left|p_{n_j} - p\right|, |\nabla \boldsymbol{\varphi}|^2\right) &= \limsup_n \left(|p_n - p|, |\nabla \boldsymbol{\varphi}|^2\right), \text{ and} \\ \lim_j p_{n_j} &= p \text{ a.e.} \end{aligned}$$

Then $\left|p_{n_j} - p\right| |\nabla \boldsymbol{\varphi}|^2 \leq 2C |\nabla \boldsymbol{\varphi}|^2$ a.e., by assumption. Therefore, by Lebesgue's Dominated Convergence Theorem,

$$\begin{aligned} \limsup_n \left(|p_n - p|, |\nabla \boldsymbol{\varphi}|^2\right) &= \lim_j \left(\left|p_{n_j} - p\right|, |\nabla \boldsymbol{\varphi}|^2\right) \\ &= \left(\lim_j \left|p_{n_j} - p\right|, |\nabla \boldsymbol{\varphi}|^2\right) = 0. \end{aligned}$$

□

Proof of Theorem 2.6.5. **Step 1:** *For $m, l \to \infty$, there is a subsequence $\{\mathbf{U}_n, S_n\}$, converging weakly in $H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega)$ to some $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega, [-1, 1])$.*

For every $m, l \in \mathbb{N}$, Proposition 2.6.4 gives existence of $\left(\mathbf{U}_m^l, S_m^l\right)$ and ensures that $-1 \leq S_m^l \leq 1$ a.e. By Lemma 2.6.3 and by assumption,

$$G_{\varepsilon,h_l}\left(\mathbf{U}_m^l, S_m^l\right) \leq G_{\varepsilon,h_l}\left(\mathbf{U}_0^l, S_0^l\right) \leq C_0,$$

independently of l, m. In particular, $G_{\varepsilon,h_n}(\mathbf{U}_n^n, S_n^n) \leq C_0$. So, by the definition of G_{ε,h_n}, the H^1-norms of $\mathbf{U}_n^n$ and S_n^n are bounded independently of n. Therefore, since H^1 is a Hilbert space, there exist subsequences, called $\{\mathbf{U}_n\}$ and $\{S_n\}$, which converge weakly in H^1 to some $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega)$.

Finally, since $H^1(\Omega)$ is a Hilbert space and $\{\varphi \in H^1(\Omega) : -1 \leq \varphi \leq 1 \text{ a.e}\} \subset H^1(\Omega)$ is a closed, convex set, it is weakly closed. Therefore, by the weak convergence in H^1 of $S_n \rightharpoonup s$, we get $-1 \leq s \leq 1$.

Below, we shall use the abbreviation h for h_n.

Step 2: $\mathbf{u}$ *solves equation* (2.6.2).

Let $\boldsymbol{\varphi} \in C^\infty(\Omega, \mathbb{R}^m)$ be fixed, $n \in \mathbb{N}$, and $h > 0$. Consider

$$\gamma\left(\left(s^2 + k_\varepsilon\right) \nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\right) + \lambda\left(\mathbf{u} - \mathbf{g}, \boldsymbol{\varphi}\right) + \frac{1}{\delta_\varepsilon}\left(\left(|\mathbf{u}|^2 - 1\right)\mathbf{u}, \boldsymbol{\varphi}\right) =: \gamma T_1 + \lambda T_2 + \frac{1}{\delta_\varepsilon} T_3.$$

Since H^1 is compactly embedded in L^p for $p < 6$, as long as the space dimension $d \leq 3$, we have $\mathbf{U}_n \to \mathbf{u}$ in $L^p(\Omega, \mathbb{R}^m)$ for $p < 6$.

We compute

$$\begin{aligned} T_1 &= \left(\left(S_{n-1}^2 + k_\varepsilon\right) \nabla \mathbf{U}_n, \nabla \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right) + \left(\left(s^2 - S_{n-1}^2\right) \nabla \mathbf{U}_n, \nabla \boldsymbol{\varphi}\right) \\ &\quad + \left(\left(s^2 + k_\varepsilon\right) \nabla\left(\mathbf{u} - \mathbf{U}_n\right), \nabla \boldsymbol{\varphi}\right) + \left(\left(S_{n-1}^2 + k_\varepsilon\right) \nabla \mathbf{U}_n, \nabla\left(\boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right)\right) \\ &=: T_{11}^n + T_{12}^n + T_{13}^n + T_{14}^n. \end{aligned}$$

Note that $\|\boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\|_{H^r(\Omega,\mathbb{R}^m)} \leq c h^{2-r} \|\nabla^2 \boldsymbol{\varphi}\|_{L^2(\Omega,\mathbb{R}^m)}$ for $0 \leq r \leq 2$.

Since $-1 \leq S_{n-1}, s \leq 1$, we have $\left|S_{n-1}^2 - s^2\right| \leq C\left|S_{n-1} - s\right| \leq C\left|S_{n-1} - s\right|^{1/2}$, whence, by Lemma 2.6.6,

$$T_{12}^n = \left(\left(s^2 - S_{n-1}^2\right)\nabla \mathbf{U}_n, \nabla \boldsymbol{\varphi}\right) \leq C\left(\left|s - S_{n-1}\right|, \left|\nabla \boldsymbol{\varphi}\right|^2\right)^{1/2} \left\|\nabla \mathbf{U}_n\right\|_{L^2\left(\Omega, \mathbb{R}^{m\times d}\right)} \xrightarrow[n\to+\infty]{h\to 0} 0.$$

Since $s \leq 1$, we know that $\left(s^2 + k_\varepsilon\right)\nabla \boldsymbol{\varphi} \in L^2\left(\Omega, \mathbb{R}^{d\times m}\right)$, so

$$T_{13}^n = \left(\nabla\left(\mathbf{u} - \mathbf{U}_n\right), \left(s^2 + k_\varepsilon\right)\nabla \boldsymbol{\varphi}\right) \xrightarrow[n\to+\infty]{h\to 0} 0,$$

by the very definition of weak convergence. And since $\left\|\boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right\|_{H^1(\Omega,\mathbb{R}^m)} \to 0$, we get

$$T_{14}^n \leq \left\|S_{n-1}^2 + k_\varepsilon\right\|_{L^\infty(\Omega)} \left\|\nabla \mathbf{U}_n\right\|_{L^2\left(\Omega,\mathbb{R}^{m\times d}\right)} \left\|\nabla\left(\boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right)\right\|_{L^2\left(\Omega,\mathbb{R}^{m\times d}\right)} \xrightarrow[n\to+\infty]{h\to 0} 0.$$

Furthermore

$$\begin{aligned} T_2 &= \left(\mathbf{U}_n - \mathbf{G}, \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right) + \left(\mathbf{U}_n - \mathbf{G}, \boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right) \\ &\quad + \left(\mathbf{r}_h(\mathbf{g}) - \mathbf{g}, \boldsymbol{\varphi}\right) + \left(\mathbf{u} - \mathbf{U}_n, \boldsymbol{\varphi}\right) \\ &=: T_{21}^n + T_{22}^n + T_{23}^n + T_{24}^n, \end{aligned}$$

with

$$\begin{aligned} T_{22}^n &\leq \left\|\mathbf{U}_n - \mathbf{G}\right\|_{L^2(\Omega,\mathbb{R}^m)} \left\|\boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right\|_{L^2(\Omega,\mathbb{R}^m)} \to 0, \\ T_{23}^n &\leq \left\|\mathbf{r}_h(\mathbf{g}) - \mathbf{g}\right\|_{L^2(\Omega,\mathbb{R}^m)} \left\|\boldsymbol{\varphi}\right\|_{L^2(\Omega,\mathbb{R}^m)} \to 0, \text{ and} \\ T_{24}^n &\leq \left\|\mathbf{u} - \mathbf{U}_n\right\|_{L^2(\Omega,\mathbb{R}^m)} \left\|\boldsymbol{\varphi}\right\|_{L^2(\Omega,\mathbb{R}^m)} \to 0. \end{aligned}$$

For the last term, we get

$$\begin{aligned} T_3 &= \left(\left(\left|\mathbf{U}_n\right|^2 - 1\right)\mathbf{U}_n, \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right) + \left(\left(\left|\mathbf{U}_n\right|^2 - 1\right)\mathbf{U}_n, \boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right) \\ &\quad + \left(\left(\left|\mathbf{u}\right|^2 - 1\right)\mathbf{u} - \left(\left|\mathbf{U}_n\right|^2 - 1\right)\mathbf{U}_n, \boldsymbol{\varphi}\right) =: T_{31}^n + T_{32}^n + T_{33}^n, \end{aligned}$$

with

$$T_{32}^n \leq \left\|\left(\left|\mathbf{U}_n\right|^2 - 1\right)\mathbf{U}_n\right\|_{L^2(\Omega,\mathbb{R}^m)} \left\|\boldsymbol{\varphi} - \boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})\right\|_{L^2(\Omega,\mathbb{R}^m)} \xrightarrow[n\to+\infty]{h\to 0} 0,$$

and

$$T_{33}^n \leq \left\|\left|\mathbf{u}\right|^2 + \mathbf{u}\cdot\mathbf{U}_n + \left|\mathbf{U}_n\right|^2 - 1\right\|_{L^2(\Omega,\mathbb{R}^m)} \left\|\mathbf{u} - \mathbf{U}_n\right\|_{L^4(\Omega,\mathbb{R}^m)} \left\|\boldsymbol{\varphi}\right\|_{L^4(\Omega,\mathbb{R}^m)} \xrightarrow[n\to+\infty]{h\to 0} 0.$$

So, putting all of the above together, we have for $n \in \mathbb{N}$ and $h > 0$ fixed,

$$\gamma\left(\left(s^2 + k_\varepsilon\right)\nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\right) + \lambda\left(\mathbf{u} - \mathbf{g}, \boldsymbol{\varphi}\right) + \frac{1}{\delta_\varepsilon}\left(\left(\left|\mathbf{u}\right|^2 - 1\right)\mathbf{u}, \boldsymbol{\varphi}\right) =: \gamma T_{13}^n + \lambda T_{21}^n + \frac{1}{\delta_\varepsilon} T_{31}^n + T^n,$$

where $\gamma T_{13}^n + \lambda T_{21}^n + \frac{1}{\delta_\varepsilon} T_{31}^n = 0$ by construction. Now, letting $n \to +\infty$ and $h \to 0$, we have $T^n \to 0$, as shown above. And by a density argument, the above is true for general $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$.

Step 3: $\nabla \mathbf{U}_n \to \nabla \mathbf{u}$ *strongly in* $L^2\left(\Omega, \mathbb{R}^{m\times d}\right)$, *as* $n \to +\infty$ *and* $h \to 0$.

Let $n \in \mathbb{N}$ and $h > 0$. Then

$$\begin{aligned}
&\gamma k_\varepsilon \left\|\nabla\left(\mathbf{u}-\mathbf{U}_n\right)\right\|_{L^2}^2 \\
&\quad \leq \gamma\left(\left(S_{n-1}^2+k_\varepsilon\right)\nabla\left(\mathbf{u}-\mathbf{U}_n\right), \nabla\left(\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)+\nabla\left(\mathbf{u}-\mathbf{R}_h(\mathbf{u})\right)\right) \\
&\quad = \gamma\left(\left(S_{n-1}^2+k_\varepsilon\right)\nabla\mathbf{u}, \nabla\left(\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)\right) \\
&\qquad -\gamma\left(\left(S_{n-1}^2+k_\varepsilon\right)\nabla\mathbf{U}_n, \nabla\left(\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)\right) \\
&\qquad -\lambda\left(\mathbf{U}_n-\mathbf{G}, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)-\frac{1}{\delta_\varepsilon}\left(\left(\left|\mathbf{U}_n\right|^2-1\right)\mathbf{U}_n, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right) \\
&\qquad +\lambda\left(\mathbf{U}_n-\mathbf{G}, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)+\frac{1}{\delta_\varepsilon}\left(\left(\left|\mathbf{U}_n\right|^2-1\right)\mathbf{U}_n, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right) \\
&\qquad +\gamma\left(\left(S_{n-1}^2+k_\varepsilon\right)\nabla\left(\mathbf{u}-\mathbf{U}_n\right), \nabla\left(\mathbf{u}-\mathbf{R}_h(\mathbf{u})\right)\right) \\
&\quad =: T_1^n+\ldots+T_7^n.
\end{aligned}$$

By construction (equation (2.6.5) with $\mathbf{W} := \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n$), the expression $T_2^n+T_3^n+T_4^n$ is zero.

$$\begin{aligned}
T_1^n &= \gamma\left(\left(S_{n-1}^2+k_\varepsilon\right)\nabla\mathbf{u}, \nabla\left(\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)\right) \\
&= \gamma\left(\left(S_{n-1}^2-s^2\right)\nabla\mathbf{u}, \nabla\left(\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)\right)+\gamma\left(\left(s^2+k_\varepsilon\right)\nabla\mathbf{u}, \nabla\left(\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)\right) \\
&\quad +\lambda\left(\mathbf{u}-\mathbf{g}, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)+\frac{1}{\delta_\varepsilon}\left(\left(\left|\mathbf{u}\right|^2-1\right)\mathbf{u}, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right) \\
&\quad -\lambda\left(\mathbf{u}-\mathbf{g}, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right)-\frac{1}{\delta_\varepsilon}\left(\left(\left|\mathbf{u}\right|^2-1\right)\mathbf{u}, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right) \\
&=: T_{11}^n+\ldots+T_{16}^n.
\end{aligned}$$

By Step 2, $T_{12}^n+T_{13}^n+T_{14}^n = 0$. Therefore

$$\gamma k_\varepsilon \left\|\nabla\left(\mathbf{u}-\mathbf{U}_n\right)\right\|_{L^2}^2 \leq T_{11}^n+T_{15}^n+T_{16}^n+T_5^n+T_6^n+T_7^n.$$

All of the above is true for any $n \in \mathbb{N}$. Now, consider the limit $n \to +\infty$ and $h \to 0$. Note that, by a density-argument,

$$\left\|\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right\|_X \leq \left\|\mathbf{R}_h(\mathbf{u})-\mathbf{u}\right\|_X+\left\|\mathbf{u}-\mathbf{U}_n\right\|_X \underset{n\to+\infty}{\overset{h\to 0}{\longrightarrow}} 0,$$

for $X = H^1$ and, by embedding, $X = L^p$ $(p < 6)$. Therefore we have, similarly to Step 2,

$$T_5^n = \lambda\left(\mathbf{U}_n-\mathbf{g}, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right) \leq \left\|\mathbf{U}_n-\mathbf{g}\right\|_{L^2(\Omega,\mathbb{R}^m)}\left\|\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right\|_{L^2(\Omega,\mathbb{R}^m)} \underset{n\to+\infty}{\overset{h\to 0}{\longrightarrow}} 0,$$

and

$$\begin{aligned}
T_6^n &= \frac{1}{\delta_\varepsilon}\left(\left(\left|\mathbf{U}_n\right|^2-1\right)\mathbf{U}_n, \mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right) \\
&\leq \frac{1}{\delta_\varepsilon}\left\|\left(\left|\mathbf{U}_n\right|^2-1\right)\mathbf{U}_n\right\|_{L^{4/3}(\Omega,\mathbb{R}^m)}\left\|\mathbf{R}_h(\mathbf{u})-\mathbf{U}_n\right\|_{L^4(\Omega,\mathbb{R}^m)} \underset{n\to+\infty}{\overset{h\to 0}{\longrightarrow}} 0,
\end{aligned}$$

and the same is true for T_{15}^n and T_{16}^n. Furthermore

$$\begin{aligned}
T_7^n &= \gamma\left(\left(S_{n-1}^2+k_\varepsilon\right)\nabla\left(\mathbf{u}-\mathbf{U}_n\right), \nabla\left(\mathbf{u}-\mathbf{R}_h(\mathbf{u})\right)\right) \\
&\leq C\left\|S_{n-1}^2+k_\varepsilon\right\|_{L^\infty(\Omega)}\left\|\nabla\left(\mathbf{u}-\mathbf{U}_n\right)\right\|_{L^2(\Omega,\mathbb{R}^m)}\left\|\nabla\left(\mathbf{u}-\mathbf{R}_h(\mathbf{u})\right)\right\|_{L^2(\Omega,\mathbb{R}^m)} \underset{n\to+\infty}{\overset{h\to 0}{\longrightarrow}} 0,
\end{aligned}$$

and finally,

$$T_{11}^n = \gamma \left\| \left(S_{n-1}^2 - s^2\right) \nabla \mathbf{u} \right\|_{L^2(\Omega,\mathbb{R}^m)} \left\| \nabla \left(\mathbf{R}_h(\mathbf{u}) - \mathbf{U}_n\right) \right\|_{L^2(\Omega,\mathbb{R}^m)} \xrightarrow[n\to+\infty]{h\to 0} 0,$$

by Lemma 2.6.6, as in Step 2.

Step 4: *s solves equation* (2.6.3)*, and* $0 \le s \le 1$.
Let $\varphi \in C^\infty(\Omega)$ be fixed, $n \in \mathbb{N}$, and $h > 0$. Set

$$2\alpha\varepsilon \left(\nabla s, \nabla\varphi\right) + \gamma \left(|\nabla \mathbf{u}|^2 s, \varphi\right) + \frac{\alpha}{2\varepsilon}\left(s-1,\varphi\right) := 2\alpha\varepsilon T_1 + \gamma T_2 + \frac{\alpha}{2\varepsilon} T_3.$$

We have

$$\begin{aligned} T_1 &= \left(\nabla S_n, \nabla \mathcal{I}_h(\varphi)\right) + \left(\nabla S_n, \nabla\left(\varphi - \mathcal{I}_h(\varphi)\right)\right) + \left(\nabla\left(s - S_n\right), \nabla\varphi\right) \\ &=: T_{11}^n + T_{12}^n + T_{13}^n, \end{aligned}$$

with $T_{12}^n, T_{13}^n \to 0$ by the strong H^1 convergence of $\mathcal{I}_h(\cdot)$ and the weak H^1 convergence of S_n, respectively, like in Step 2.

Also,

$$\begin{aligned} T_2 &= \left(|\nabla \mathbf{U}_n|^2 S_n, \mathcal{I}_h(\varphi)\right) + \left(|\nabla \mathbf{U}_n|^2 S_n, \varphi - \mathcal{I}_h(\varphi)\right) \\ &\quad + \left(\left(|\nabla \mathbf{u}|^2 - |\nabla \mathbf{U}_n|^2\right) S_n, \varphi\right) + \left(|\nabla \mathbf{u}|^2 \left(s - S_n\right), \varphi\right) \\ &=: T_{11}^n + T_{12}^n + T_{13}^n + T_{14}^n. \end{aligned}$$

Since $\varphi \in C^\infty(\Omega)$, we have $\mathcal{I}_h(\varphi) \to \varphi$ in $L^\infty(\Omega)$ for $h \to 0$, leading to

$$T_{12}^n \le \|\nabla \mathbf{U}_n\|_{L^2\left(\Omega,\mathbb{R}^{m\times d}\right)}^2 \|S_n\|_{L^\infty(\Omega)} \|\varphi - \mathcal{I}_h(\varphi)\|_{L^\infty(\Omega)} \xrightarrow[n\to+\infty]{h\to 0} 0,$$

By Step 3, we conclude

$$T_{13}^n \le C \|\nabla \mathbf{u} + \nabla \mathbf{U}_n\|_{L^2\left(\Omega,\mathbb{R}^{m\times d}\right)} \|\nabla \mathbf{u} - \nabla \mathbf{U}_n\|_{L^2\left(\Omega,\mathbb{R}^{m\times d}\right)} \|S_n\|_{L^\infty(\Omega)} \|\varphi\|_{L^2(\Omega)} \xrightarrow[n\to+\infty]{h\to 0} 0,$$

and

$$T_{14}^n \le C \left(|s - S_n|, |\nabla \mathbf{u}|^2\right) \|\varphi\|_{L^\infty(\Omega)} \xrightarrow[n\to+\infty]{h\to 0} 0,$$

by Lemma 2.6.6, like in Step 2.

Finally

$$\begin{aligned} T_3 &= \left(S_n - 1, \mathcal{I}_h(\varphi)\right)_h + \left(S_n - 1, \mathcal{I}_h(\varphi)\right) - \left(S_n - 1, \mathcal{I}_h(\varphi)\right)_h \\ &\quad + \left(S_n - 1, \varphi - \mathcal{I}_h(\varphi)\right) + \left(s - S_n, \varphi\right) \\ &=: T_{31}^n + \ldots + T_{35}^n, \end{aligned}$$

with $|T_{32}^n - T_{33}^n| \le Ch \|\nabla S_n\|_{L^2(\Omega)} \|\mathcal{I}_h(\varphi)\|_{L^2(\Omega)} \to 0$, and $T_{34}^n, T_{35}^n \to 0$ by the strong L^p convergence of $\mathcal{I}_h(\cdot)$ and S_n, respectively.

So, putting all of the above together, we have for $n \in \mathbb{N}$ and $h > 0$ fixed,

$$2\alpha\varepsilon\left(\nabla s, \nabla\varphi\right) + \gamma\left(|\nabla \mathbf{u}|^2 s, \varphi\right) + \frac{\alpha}{2\varepsilon}\left(s-1, \varphi\right) =: 2\alpha\varepsilon T_{13}^n + \gamma T_{21}^n + \frac{\alpha}{2\varepsilon} T_{31}^n + T^n,$$

where $2\alpha\varepsilon T_{13}^n + \gamma T_{21}^n + \frac{\alpha}{2\varepsilon} T_{31}^n = 0$ by construction. Now, letting $n \to +\infty$ and $h \to 0$, we get $T^n \to 0$, as shown above.

So we have shown that s solves equation (2.6.3) for test functions $\varphi \in C^\infty(\Omega)$. By a density argument, this is also true for general $\varphi \in H^1(\Omega)$. Therefore s is a critical point of G_ε. And since replacing s pointwise by $0 \vee s \wedge 1$ would only decrease every term of this energy, $0 \leq s \leq 1$ follows.

This concludes the proof. □

Remark 2.6.7. *For $d \leq 2$, one can also get $\nabla S_n \to \nabla s$ strongly in $L^2(\Omega; \mathbb{R}^m)$, with an argument similar to Step 3, using the equations for S_n and s and a test function $R_h(s) - S_n$. It breaks down for $d \geq 3$ because of the lack of L^∞-stability of the Ritz projection.*

2.7 Detour: Regularised Penalisation & Splitting

Before being made aware of the preprint [25], we lacked the idea of using Dominated Convergence (Lemma 2.6.6) — obvious though this idea looks in hindsight —, and therefore couldn't get strong convergence for $\nabla \mathbf{u}_n$ in L^2. We therefore added another regularising term $\frac{\eta_\varepsilon}{p}\int_\Omega |\Delta s|^p \mathrm{d}\mathbf{x}$. This does allow identification of limits and does not destroy Γ-convergence if η_ε is properly scaled with respect to ε, and we give an abbreviated account of this approach in this section, because we believe it contains some interesting mathematics. However, the practical disadvantages of this approach are obvious: The need for a Hermite finite element space $V_h(\Omega) \subset W^{2,p}(\Omega)$ (or a mixed formulation which further complicates the analysis) and no maximum principle for s.

2.7.1 Γ-Convergence

Let $\Omega \subset \mathbb{R}^d$, γ, α, λ be fixed positive constants, $\varepsilon, \delta_\varepsilon > 0$, $k_\varepsilon, \eta_\varepsilon \geq 0$, $\mathbf{g} \in L^\infty(\Omega, \mathbb{S}^{m-1})$, $1 < p < +\infty$, and $G_\varepsilon, G : L^2(\Omega, \mathbb{R}^m) \times L^2(\Omega) \to [0, +\infty]$ be defined by

$$G_\varepsilon(\mathbf{u}, s) := \begin{cases} \dfrac{\gamma}{2}\displaystyle\int_\Omega \left(s^2 + k_\varepsilon\right)|\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x} & \text{if } \mathbf{u} \in H^1(\Omega, \mathbb{R}^m), \\ \quad + \alpha \displaystyle\int_\Omega \left(\varepsilon |\nabla s|^2 + \frac{(1-s)^2}{4\varepsilon}\right)\mathrm{d}\mathbf{x} & s \in W^{2,p}(\Omega, [0,1]), \\ \quad + \dfrac{1}{\delta_\varepsilon}\displaystyle\int_\Omega \left(|\mathbf{u}|^2 - 1\right)^2 \mathrm{d}\mathbf{x} + \frac{\eta_\varepsilon}{p}\int_\Omega |\Delta s|^p \mathrm{d}\mathbf{x} & \\ +\infty, & \text{otherwise,} \end{cases}$$

and

$$G(\mathbf{u}, s) := \begin{cases} \dfrac{\gamma}{2}\displaystyle\int_\Omega |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \alpha \mathcal{H}^{d-1}(S_{\mathbf{u}}) + \frac{\lambda}{2}\int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x} & \text{if } \mathbf{u} \in GSBV(\Omega, \mathbb{S}^{m-1}) \\ & \text{and } s = 1 \text{ a.e.} \\ +\infty, & \text{otherwise.} \end{cases}$$

We shall use the following assumption on account of the term involving second derivatives. In the case of the original Ambrosio-Tortorelli energy, no such assumption is necessary, and we expect that even in our case a density result like [39, 41] and a diagonal argument should allow to overcome this restriction. But we do not address this problem here.

Assumption 2.7.1. *Let $S_{\mathbf{u}}$ consists of a finite number of disjoint C^2 manifolds of dimension $(d-1)$.*

Theorem 2.7.2. *If $\Omega \subset \mathbb{R}^d$ is open and bounded with Lipschitz boundary, $\delta_\varepsilon \underset{\varepsilon\to 0}{\longrightarrow} 0$, $k_\varepsilon = o(\varepsilon)$, $k_\varepsilon = o(\delta_\varepsilon)$, $\eta_\varepsilon = o(k_\varepsilon^{p-1}\varepsilon^p)$, and $S_{\mathbf{u}}$ fulfils Assumption 2.7.1, then $G_\varepsilon \xrightarrow[\varepsilon\to 0]{\Gamma} G$ in $L^2(\Omega,\mathbb{R}^m)\times L^2(\Omega)$.*

Moreover, there exists a solution $\{\mathbf{u}_\varepsilon, s_\varepsilon\}$ of the minimum problem

$$m_\varepsilon = \inf_{\substack{\mathbf{u}\in H^1(\Omega,\mathbb{R}^m),\\ s\in W^{2,p}(\Omega,[0,1])}} G_\varepsilon(\mathbf{u},s)$$

with $\|\mathbf{u}_\varepsilon\|_{L^\infty} \le C$, and every cluster point of $\{\mathbf{u}_\varepsilon\}$ is a solution of the minimum problem

$$m = \inf_{\mathbf{u}\in GSBV(\Omega,\mathbb{S}^{m-1})} G(\mathbf{u},1),$$

and $m_\varepsilon \to m$ as $\varepsilon \to 0^+$.

Sketch of Proof. The lim inf inequality follows like in Section 2.5

For the limsup inequality, let $\mathbf{u} \in SBV(\Omega,\mathbb{R}^m)\cap L^\infty(\Omega,\mathbb{R}^m)$ with $\nabla\mathbf{u} \in L^2\big(\Omega,\mathbb{R}^{d\times m}\big)$, $|\mathbf{u}|^2 = 1$ a.e., $S_{\mathbf{u}}$ essentially closed in Ω, and $s = 1$ a.e. We then have to construct $(\mathbf{u}_\varepsilon, s_\varepsilon)$ that converge in $L^2(\Omega,\mathbb{R}^m)\times L^2(\Omega)$ to $(\mathbf{u},s)$, such that

$$\limsup_{\varepsilon\to 0^+} G_\varepsilon\left(\mathbf{u}_\varepsilon, s_\varepsilon\right) \le G(\mathbf{u},s)$$

for any positive sequence ε converging to zero.

To this end, we use a smoothed version of the profile used in Section 2.5, to ensure $s_\varepsilon \in W^{2,p}(\Omega)$: Choose a positive sequence b_ε, such that $b_\varepsilon = o(\varepsilon)$, $b_\varepsilon = o(\delta_\varepsilon)$, and $k_\varepsilon = o(b_\varepsilon)$. For any $b > 0$, set $S_b := \{\mathbf{x}\in\Omega : d(\mathbf{x}) < b\}$. Thanks to (2.5.2), $|S_b| = O(b)$. For $t \ge 2b_\varepsilon$, let

$$\begin{aligned}
\sigma_\varepsilon(t) &:= 1 - \exp\left(-\frac{t-2b_\varepsilon}{2\varepsilon}\right), \text{ so that}\\
\sigma_\varepsilon'(t) &= \frac{1}{2\varepsilon}\exp\left(-\frac{t-2b_\varepsilon}{2\varepsilon}\right), \text{ and}\\
\sigma_\varepsilon''(t) &= -\frac{1}{4\varepsilon^2}\exp\left(-\frac{t-2b_\varepsilon}{2\varepsilon}\right).
\end{aligned}$$

We would now like to set (c.f. Figure 2.2, left)

$$s_\varepsilon(\mathbf{x}) := \begin{cases} 0 & \text{if } \mathbf{x}\in S_{2b_\varepsilon},\\ \sigma_\varepsilon(d(\mathbf{x})) & \text{if } \mathbf{x}\in S_{2b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}}\setminus S_{2b_\varepsilon},\\ 1-c_\varepsilon & \text{if } \mathbf{x}\in\Omega\setminus S_{2b_\varepsilon+2\varepsilon\ln\frac{1}{\varepsilon}}. \end{cases} \tag{2.7.1}$$

Note that $0 < 2\varepsilon \ln \frac{1}{\varepsilon} \to 0^+$, and $\varepsilon = o\big(2\varepsilon \ln \frac{1}{\varepsilon}\big)$. Unfortunately, for this choice, $s_\varepsilon \notin W^{2,p}(\Omega)$. So we have to smooth out the two kinks in s_ε. This can be done by extending/replacing σ_ε in $t \in [b_\varepsilon, 3b_\varepsilon]$ and $t \in \big[b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}, 3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\big]$ with polynomials $p_\varepsilon, q_\varepsilon$ of degree 3 such that $p_\varepsilon(b_\varepsilon) = p'_\varepsilon(b_\varepsilon) = q_\varepsilon\big(3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\big) = q'_\varepsilon\big(3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\big) = 0$, $p_\varepsilon(3b_\varepsilon) = \sigma_\varepsilon(3b_\varepsilon)$, $p'_\varepsilon(3b_\varepsilon) = \sigma'_\varepsilon(3b_\varepsilon)$, $q_\varepsilon\big(b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\big) = \sigma_\varepsilon\big(b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\big)$, $q'_\varepsilon\big(b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\big) = \sigma'_\varepsilon\big(b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\big)$. A simple calculation shows that this is possible while keeping

$$p_\varepsilon, q_\varepsilon = O(1)\,, \qquad p'_\varepsilon, q'_\varepsilon = O\left(\frac{1}{\varepsilon}\right), \qquad p''_\varepsilon, q''_\varepsilon = O\left(\frac{1}{\varepsilon b_\varepsilon}\right). \tag{2.7.2}$$

Now, set (c.f. Figure 2.2, right)

$$s_\varepsilon(\mathbf{x}) := \begin{cases} 0 & \text{if } \mathbf{x} \in S_{b_\varepsilon}, \\ p_\varepsilon(d(\mathbf{x})) & \text{if } \mathbf{x} \in S_{3b_\varepsilon} \setminus b_\varepsilon, \\ \sigma_\varepsilon(d(\mathbf{x})) & \text{if } \mathbf{x} \in S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \setminus S_{3b_\varepsilon}, \\ q_\varepsilon(d(\mathbf{x})) & \text{if } \mathbf{x} \in S_{3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \setminus S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}}, \\ 1 - c_\varepsilon & \text{if } \mathbf{x} \in \Omega \setminus S_{3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}}, \end{cases} \tag{2.7.3}$$

and

$$\mathbf{u}_\varepsilon(\mathbf{x}) := \mathbf{u}(\mathbf{x}) \min\left\{\frac{d(\mathbf{x})}{b_\varepsilon}, 1\right\}.$$

By the above, indeed $s_\varepsilon \in W^{2,p}(\Omega)$.

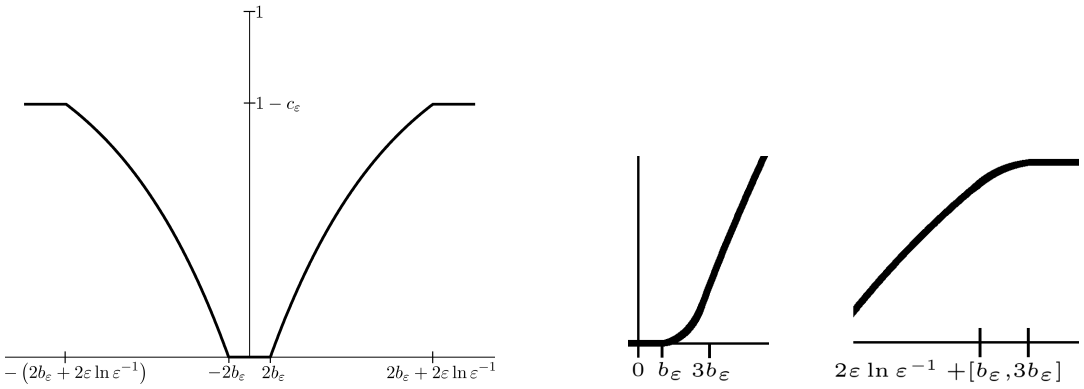

Figure 2.2: Sketch of $s_\varepsilon(\mathbf{x})$ in the case $S_\mathbf{u} = \{0\}$, and $d = 1$; original (left) and two crops of the smoothed version (right).

The calculations are now the same as in Section 2.5, except for the terms with coefficients α and $\frac{\eta_\varepsilon}{p}$ in the energy. For the first of these, we have

$$\begin{aligned} &\int_\Omega \left(\varepsilon |\nabla s_\varepsilon|^2 + \frac{(1 - s_\varepsilon)^2}{4\varepsilon}\right) \mathrm{d}\mathbf{x} \\ &\quad = \left[\int_{S_{3b_\varepsilon}} + \int_{S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \setminus S_{3b_\varepsilon}} + \int_{\Omega \setminus S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}}}\right] \left(\varepsilon \, |\nabla s_\varepsilon|^2 + \frac{(1 - s_\varepsilon)^2}{4\varepsilon}\right) \mathrm{d}\mathbf{x} \\ &\quad =: I_\varepsilon + II_\varepsilon + III_\varepsilon. \end{aligned}$$

Apart from the obvious changes, the calculations for the term II_ε are again the same as in Section 2.5. What is left to show is $I_\varepsilon, III_\varepsilon \to 0$. Indeed, by the construction of s_ε, $|S_{b_\varepsilon}| = O(b_\varepsilon)$, $b_\varepsilon = o(\varepsilon)$, and (2.7.2),

$$\begin{aligned} I_\varepsilon &\leq \int_{S_{b_\varepsilon}} \frac{1}{4\varepsilon} \mathrm{d}\mathbf{x} + \int_{S_{3b_\varepsilon} \setminus S_{b_\varepsilon}} \left(\varepsilon \left(p_\varepsilon'(\mathbf{x})\right)^2 + \frac{\left(1 - p_\varepsilon(\mathbf{x})\right)^2}{4\varepsilon} \right) \mathrm{d}\mathbf{x} \\ &\leq C \frac{b_\varepsilon}{\varepsilon} \to 0. \end{aligned}$$

And similarly,

$$\begin{aligned} III_\varepsilon &= \int_{S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \setminus S_{3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}}} \left(\varepsilon \left(q_\varepsilon'(\mathbf{x})\right)^2 + \frac{\left(1 - q_\varepsilon(\mathbf{x})\right)^2}{4\varepsilon} \right) \mathrm{d}\mathbf{x} + \int_{\Omega \setminus S_{3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}}} \frac{\varepsilon^2}{4\varepsilon} \mathrm{d}\mathbf{x} \\ &\leq C \left(\frac{b_\varepsilon}{\varepsilon} + \frac{\varepsilon^2}{\varepsilon} \right) \to 0, \end{aligned}$$

so, indeed,

$$\limsup_{\varepsilon \to 0^+} \int_\Omega \left(\varepsilon \left|\nabla s_\varepsilon\right|^2 + \frac{\left(1 - s_\varepsilon\right)^2}{4\varepsilon} \right) \mathrm{d}\mathbf{x} \leq \mathcal{H}^{d-1}(S_\mathbf{u}), \tag{2.7.4}$$

And for the term with coefficient $\frac{\eta_\varepsilon}{p}$, by construction,

$$|\Delta s_\varepsilon(\mathbf{x})| = \begin{cases} |p_\varepsilon''(d(\mathbf{x}))| \left|\nabla d(\mathbf{x})\right|^2 + |p_\varepsilon'(d(\mathbf{x}))| \left|\Delta d(\mathbf{x})\right| & \text{if } \mathbf{x} \in S_{3b_\varepsilon} \setminus b_\varepsilon, \\ |\sigma_\varepsilon''(d(\mathbf{x}))| \left|\nabla d(\mathbf{x})\right|^2 + |\sigma_\varepsilon'(d(\mathbf{x}))| \left|\Delta d(\mathbf{x})\right| & \text{if } \mathbf{x} \in S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \setminus S_{3b_\varepsilon}, \\ |q_\varepsilon''(d(\mathbf{x}))| \left|\nabla d(\mathbf{x})\right|^2 + |q_\varepsilon'(d(\mathbf{x}))| \left|\Delta d(\mathbf{x})\right| & \text{if } \mathbf{x} \in S_{3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \setminus S_{b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}}, \\ 0 & \text{otherwise.} \end{cases}$$

We know that $|\nabla d(\mathbf{x})| = 1$ a.e. and $S_\mathbf{u}$ consists of a finite number of disjoint C^2 manifolds of dimension $d-1$. So for ε small enough no two of the $\left(3b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}\right)$ neighbourhoods of the disjoint components of $S_\mathbf{u}$ intersect, and $|\Delta d(\mathbf{x})| = O(d^{-1}(\mathbf{x})) = O(b_\varepsilon) = o(k_\varepsilon)$. Therefore,

$$\begin{aligned} &\frac{\eta_\varepsilon}{p} \int_\Omega |\Delta s_\varepsilon|^p \, \mathrm{d}\mathbf{x} \\ &\quad \leq C\eta_\varepsilon \left(\frac{b_\varepsilon}{(\varepsilon b_\varepsilon)^p} + \int_{2b_\varepsilon}^{2b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \left(\frac{1}{\varepsilon^{2p}} + \frac{1}{t^p \varepsilon^p} \right) \exp\left(-p \frac{t - 2b_\varepsilon}{2\varepsilon} \right) \mathrm{d}t \right) \\ &\quad \leq C\eta_\varepsilon \left(\frac{1}{\varepsilon^p b_\varepsilon^{p-1}} + \frac{1}{\varepsilon^p} \int_{2b_\varepsilon}^{2b_\varepsilon + 2\varepsilon \ln \frac{1}{\varepsilon}} \frac{1}{t^p} \mathrm{d}t \right) \\ &\quad \leq C \frac{\eta_\varepsilon}{\varepsilon^p b_\varepsilon^{p-1}} \leq C \frac{\eta_\varepsilon}{\varepsilon^p k_\varepsilon^{p-1}} \to 0. \end{aligned} \tag{2.7.5}$$

Putting this together proves the lim sup inequality.

Finally, convergence of minimisers again follows like in Section 2.5. □

2.7.2 Continuous Algorithm and Analysis

Let $\Omega \subset \mathbb{R}^d$ be a convex, bounded Lipschitz domain, $\mathbf{g} : \Omega \to \mathbb{S}^{m-1}$ the chromaticity component of a given image, and $1 < p < +\infty$ ($p > d$ will later be required to get the right Sobolev embeddings). For $\mathbf{u}, \mathbf{g} \in H^1(\Omega, \mathbb{R}^m)$, $s \in W^{2,p}(\Omega)$, and $0 < \varepsilon, k_\varepsilon, \delta_\varepsilon, \eta_\varepsilon \ll 1$, we want to minimise the following vector valued Ambrosio-Tortorelli-Ginzburg-Landau energy using a splitting strategy

$$\begin{aligned} G_\varepsilon(\mathbf{u}, s) \;=\; & \frac{\gamma}{2}\int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_\Omega |\mathbf{u} - \mathbf{g}|^2 \,\mathrm{d}\mathbf{x} + \frac{\eta_\varepsilon}{p}\int_\Omega |\Delta s|^p \,\mathrm{d}\mathbf{x} \\ & + \alpha \int_\Omega \varepsilon \,|\nabla s|^2 + \frac{1}{4\varepsilon}(1-s)^2 \mathrm{d}\mathbf{x} + \frac{1}{\delta_\varepsilon}\int_\Omega \left(|\mathbf{u}|^2 - 1\right)^2 \mathrm{d}\mathbf{x}. \end{aligned} \tag{2.7.6}$$

We shall always assume γ, α, λ to be fixed and positive. The reason for the term $\frac{\eta_\varepsilon}{p}\int_\Omega |\Delta s|^p \,\mathrm{d}\mathbf{x}$ is that for $p > d$ it allows us, by embedding, to bound the $C^{0,1}$ norm of iterates s_{n-1}^2, independently of n. Since $\operatorname{div}\left(\left(s_{n-1}^2 + k_\varepsilon\right)\nabla \mathbf{u}_n\right)$ is the highest order term in (2.7.9) (weak formulation), this, together with a standard elliptic regularity result, allows us to bound the $H^2(\Omega, \mathbb{R}^m)$ norm of $\mathbf{u}$ by the initial energy. With this we can then pass to the limit ($n \to \infty$) in the term $\left(|\nabla \mathbf{u}_n|^2 s_n, \boldsymbol{\varphi}\right)$.

Definition 2.7.3. *A tuple $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{R}^m) \times W^{2,p}(\Omega)$ is called a* weak solution *of the problem $\inf G_\varepsilon$, if and only if*

(1) *for all $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$, there holds*

$$\gamma\left(\left(s^2 + k_\varepsilon\right)\nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\right) + \lambda\left(\mathbf{u} - \mathbf{g}, \boldsymbol{\varphi}\right) + \frac{4}{\delta_\varepsilon}\left(\left(|\mathbf{u}|^2 - 1\right)\mathbf{u}, \boldsymbol{\varphi}\right) = 0, \text{ and} \tag{2.7.7}$$

(2) *for all $\varphi \in W^{2,p}(\Omega)$, there holds*

$$2\alpha\varepsilon\left(\nabla s, \nabla \varphi\right) + \gamma\left(|\nabla \mathbf{u}|^2 s, \varphi\right) + \frac{\alpha}{2\varepsilon}\left(s - 1, \varphi\right) + \eta_\varepsilon\left(|\Delta s|^{p-2}\Delta s, \Delta \varphi\right) = 0. \tag{2.7.8}$$

Algorithm 2.7.4. *Let $\mathbf{g}$ and $0 < \varepsilon, k_\varepsilon, \delta_\varepsilon, \eta_\varepsilon$ be given, and choose initial values $\mathbf{u}_0 \in H^1(\Omega, \mathbb{R}^m)$, and $s_0 \in C^{0,1}(\Omega)$ (typically, $\mathbf{u}_0 := \mathbf{g}$ and $s_0 \equiv c$ for some constant $0 \le c \le 1$). For $n = 1, \ldots$ until convergence, do*

(1) *Minimise $G_\varepsilon(\mathbf{u}, s_{n-1})$ for $\mathbf{u} \in H^1(\Omega, \mathbb{R}^m)$, and call the solution $\mathbf{u}_n$, i.e., for all $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$ solve*

$$\gamma\left(\left(s_{n-1}^2 + k_\varepsilon\right)\nabla \mathbf{u}_n, \nabla \boldsymbol{\varphi}\right) + \lambda\left(\mathbf{u}_n - \mathbf{g}, \boldsymbol{\varphi}\right) + \frac{4}{\delta_\varepsilon}\left(\left(|\mathbf{u}_n|^2 - 1\right)\mathbf{u}_n, \boldsymbol{\varphi}\right) = 0. \tag{2.7.9}$$

(2) *Minimise $G_\varepsilon(\mathbf{u}_n, s)$ for $s \in W^{2,p}(\Omega)$ and call the solution s_n, i.e., for all $\varphi \in W^{2,p}(\Omega)$ solve*

$$\begin{aligned} 2\alpha\varepsilon\left(\nabla s_n, \nabla \varphi\right) + \gamma\left(|\nabla \mathbf{u}_n|^2 s_n, \varphi\right) + \frac{\alpha}{2\varepsilon}\left(s_n - 1, \varphi\right) \\ + \eta_\varepsilon\left(|\Delta s_n|^{p-2}\,\Delta s_n, \Delta \varphi\right) = 0. \end{aligned} \tag{2.7.10}$$

Remark 2.7.5. *Algorithm 2.7.4 is energy-decreasing, since for any $n \geq 0$ it ensures*

$$G_\varepsilon(\mathbf{u}_{n+1}, s_{n+1}) \leq G_\varepsilon(\mathbf{u}_{n+1}, s_n) \leq G_\varepsilon(\mathbf{u}_n, s_n)\,.$$

The following result follows by standard coercivity and convexity arguments for G_ε, and elliptic regularity theory (see e.g. [64, Section 8.4]).

Proposition 2.7.6. *Let $k_\varepsilon, \delta_\varepsilon, \gamma > 0$. Then there exists $\mathbf{u} \in H^1(\Omega, \mathbb{R}^m)$, such that equation (2.7.7) holds for all $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$.*

If either $\Omega \subset \mathbb{R}^2$ is a bounded, convex polygon, or $\Omega \subset \mathbb{R}^d$ ($d > 2$) is $C^{1,1}$, and $s \in C^{0,1}(\Omega)$, then $\mathbf{u} \in H^2(\Omega, \mathbb{R}^m)$. Moreover, there exists a constant $C > 0$, such that

$$\|\mathbf{u}\|_{H^2} \leq C\left(\|\mathbf{u}\|_{H^1} + \left\|\lambda(\mathbf{u}-\mathbf{g}) + \frac{4}{\delta_\varepsilon}\left(|\mathbf{u}|^2 - 1\right)\mathbf{u}\right\|\right),$$

where $C = C\left(\partial\Omega, \|s^2\|_{C^{0,1}(\overline{\Omega})}, k_\varepsilon, \gamma\right) > 0$.

The following technical results are well-known.

Lemma 2.7.7. *Let $p \in (1, \infty)$. For $\gamma \geq 0$, there exist positive constants C_1, C_2, such that for all $\boldsymbol{\zeta}, \boldsymbol{\eta} \in \mathbb{R}^d$ there holds*

$$\text{(i)} \qquad \left||\boldsymbol{\zeta}|^{p-2}\boldsymbol{\zeta} - |\boldsymbol{\eta}|^{p-2}\boldsymbol{\eta}\right| \leq C_1\big(|\boldsymbol{\zeta}| + |\boldsymbol{\eta}|\big)^{p-2+\gamma}|\boldsymbol{\zeta} - \boldsymbol{\eta}|^{1-\gamma}\,,$$

$$\text{(ii)} \qquad \left\langle |\boldsymbol{\zeta}|^{p-2}\boldsymbol{\zeta} - |\boldsymbol{\eta}|^{p-2}\boldsymbol{\eta}, \boldsymbol{\zeta} - \boldsymbol{\eta}\right\rangle \geq C_2\big(|\boldsymbol{\zeta}| + |\boldsymbol{\eta}|\big)^{p-2-\gamma}|\boldsymbol{\zeta} - \boldsymbol{\eta}|^{2+\gamma}\,.$$

Browder and Minty's Theorem (using Lemma 2.7.7(ii)), together with strict convexity of the functional with respect to second derivatives of s, ensures the following proposition.

Proposition 2.7.8. *Let $\mathbf{u} \in H^2(\Omega, \mathbb{R}^m)$, $p \in (1, \infty)$, and $\varepsilon, \eta_\varepsilon > 0$. Then there exists a unique $s \in W^{2,p}(\Omega)$ such that for all $\varphi \in W^{2,p}(\Omega)$, equation (2.7.8) holds.*

Theorem 2.7.9. *Let either $\Omega \subset \mathbb{R}^2$ be a bounded, convex polygon, or $\Omega \subset \mathbb{R}^d$ ($d > 2$) be in $C^{1,1}$. Let furthermore $G_\varepsilon(\mathbf{u}_0, s_0) \leq C_0 < +\infty$ and $d < p < +\infty$. Then the sequence $\{\mathbf{u}_n, s_n\}$ constructed by Algorithm 2.7.4 converges weakly (up to subsequences in $\mathbf{u}$) in $H^2(\Omega, \mathbb{R}^m) \times W^{2,p}(\Omega)$ to some $(\mathbf{u}, s) \in H^2(\Omega, \mathbb{R}^m) \times W^{2,p}(\Omega)$, which is a weak solution as in Definition 2.7.3.*

Proof. **Step 1:** *For $n \to \infty$, $\{\mathbf{u}_n, s_n\}$ converges weakly (up to subsequences in $\mathbf{u}$) in $H^2(\Omega, \mathbb{R}^m) \times W^{2,p}(\Omega)$ to some $(\mathbf{u}, s) \in H^2(\Omega, \mathbb{R}^m) \times W^{2,p}(\Omega)$*

Thanks to Proposition 2.7.8, we know that $s_{n-1} \in W^{2,p}(\Omega)$, which is compactly embedded in $C^1\left(\overline{\Omega}\right)$ (for Ω Lipschitz and $p > d$), so together with Proposition 2.7.6 we have that Algorithm 2.7.4 makes sense. Furthermore,

$$\|\mathbf{u}_n\|_{H^2} \leq C\left(\|\mathbf{u}_n\|_{H^1} + \left\|\lambda(\mathbf{u}_n - \mathbf{g}) + \frac{4}{\delta_\varepsilon}\left(|\mathbf{u}_n|^2 - 1\right)\mathbf{u}_n\right\|\right), \tag{2.7.11}$$

with $C = C\left(\partial\Omega, \left\|s_{n-1}^2\right\|_{C^{0,1}(\overline{\Omega})}, k_\varepsilon, \gamma\right) > 0$. And since $p > d$, there is a constant $\widetilde{C} \geq 0$, such that

$$\left\|s_{n-1}^2\right\|_{C^{0,1}(\overline{\Omega})} \leq \widetilde{C}\,\|s_{n-1}\|_{W^{2,p}}^p\,,$$

which in turn is bounded by the initial energy (by Remark 2.7.5). So, all of the right hand side of (2.7.11) is bounded by some constant depending only on the initial data. Hence we can extract a subsequence, which by the above embedding has the following convergence properties for $n \to \infty$,

$$\begin{aligned}
(\mathbf{u}_n, s_n) &\rightharpoonup (\mathbf{u}, s) \quad \text{in } H^2(\Omega, \mathbb{R}^m) \times W^{2,p}(\Omega), \\
(\mathbf{u}_n, s_n) &\to (\mathbf{u}, s) \quad \text{in } C^0\big(\overline{\Omega}, \mathbb{R}^m\big) \times C^1\left(\overline{\Omega}\right), \\
\mathbf{u}_n &\to \mathbf{u} \quad \text{in } W^{1,q}(\Omega, \mathbb{R}^m) \text{ if } \begin{cases} 1 < q < \infty & \text{for } d = 2, \\ 1 < q < \frac{2d}{d-2} & \text{for } d > 2, \end{cases} \\
|\Delta s_n|^{p-2} \Delta s_n &\rightharpoonup b \quad \text{in } L^{\frac{p}{p-1}}(\Omega) \text{ for some } b \in L^{\frac{p}{p-1}}(\Omega).
\end{aligned} \tag{2.7.12}$$

In particular, $s_n \to s$ (strongly) in $L^\infty(\Omega)$, by embedding. Also, $\mathbf{u}_n \to \mathbf{u}$ (strongly) in $H^1(\Omega, \mathbb{R}^m)$, which is actually enough to pass to the limit (we do not need weak $H^2(\Omega, \mathbb{R}^m)$ convergence and indeed do not have it in the discrete case).

Step 2: $(\mathbf{u}, s)$ *is a weak solution of the problem* $\inf G_\varepsilon(\cdot, \cdot)$

For equation (2.7.9), we want to show that the following expression vanishes for $n \to \infty$

$$\gamma\Big(\left(s_n^2 + k_\varepsilon\right) \nabla \mathbf{u}_n - \left(s^2 + k_\varepsilon\right) \nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\Big) + \frac{4}{\delta_\varepsilon} \left(\left(|\mathbf{u}_n|^2 - 1\right)\mathbf{u}_n - \left(|\mathbf{u}|^2 - 1\right) \mathbf{u}, \boldsymbol{\varphi}\right) =: \gamma I + \frac{4}{\delta_\varepsilon} II,$$

the fidelity term $\lambda\left(\mathbf{u}_n - \mathbf{g} - (\mathbf{u} - \mathbf{g})\right)$ being trivial. We can assume $\boldsymbol{\varphi} \in C^\infty(\Omega, \mathbb{R}^m)$ and then have

$$\begin{aligned}
I &= k_\varepsilon\Big(\left(\nabla \left(\mathbf{u}_n - \mathbf{u}\right)\right), \nabla \boldsymbol{\varphi}\Big) + \Big(s_n^2 \nabla \mathbf{u}_n - s^2 \nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\Big) \\
&\leq C \left\| \nabla \left(\mathbf{u}_n - \mathbf{u}\right)\right\| \left\| \nabla \boldsymbol{\varphi}\right\| + C \left\| s_n^2 \nabla \mathbf{u}_n - s^2 \nabla \mathbf{u}\right\|_{L^1} \left\| \nabla \boldsymbol{\varphi}\right\|_{L^\infty} \\
&\leq C \left\| \mathbf{u}_n - \mathbf{u}\right\|_{H^1} \left\| \nabla \boldsymbol{\varphi}\right\| + C \left\| \left(s_n^2 - s^2\right) \nabla \mathbf{u}_n + s^2 \nabla \left(\mathbf{u}_n - \mathbf{u}\right)\right\|_{L^1} \left\| \nabla \boldsymbol{\varphi}\right\|_{L^\infty} \\
&\leq C \left\| \mathbf{u}_n - \mathbf{u}\right\|_{H^1} \left\| \nabla \boldsymbol{\varphi}\right\| + C \left\| s_n - s\right\|_{L^\infty} \left\| s_n + s\right\|_{L^\infty} \left\| \nabla \mathbf{u}_n\right\|_{L^1} \left\| \nabla \boldsymbol{\varphi}\right\|_{L^\infty} \\
&\quad + C \left\| s\right\|_{L^\infty}^2 \left\| \nabla \left(\mathbf{u}_n - \mathbf{u}\right)\right\|_{L^1} \left\| \nabla \boldsymbol{\varphi}\right\|_{L^\infty},
\end{aligned}$$

every single term of which tends to zero by (2.7.12). Furthermore, again by (2.7.12),

$$\begin{aligned}
II &= \Big(\left(|\mathbf{u}_n|^2 - |\mathbf{u}|^2 \right) \mathbf{u}_n, \boldsymbol{\varphi}\Big) + \Big(\left(\mathbf{u}_n - \mathbf{u}\right) |\mathbf{u}|^2, \boldsymbol{\varphi}\Big) + \left(\mathbf{u} - \mathbf{u}_n, \boldsymbol{\varphi}\right) \\
&\leq C \|\boldsymbol{\varphi}\| \left(\left\| \mathbf{u}_n - \mathbf{u}\right\|_{L^6} \left\| \mathbf{u}_n + \mathbf{u}\right\|_{L^6} \left\| \mathbf{u}_n\right\|_{L^6} + \left\| \mathbf{u}_n - \mathbf{u}\right\|_{L^6} \left\| \mathbf{u}\right\|_{L^6}^2 + \left\| \mathbf{u}_n - \mathbf{u}\right\| \right),
\end{aligned}$$

every single term of which again tends to zero.

For equation (2.7.10) we easily see that $\lim_{n\to\infty} |\left(\nabla \left(s_n - s\right), \nabla \varphi\right)| + |\left(s_n - s, \varphi\right)| = 0$. For the first of the remaining terms, we get, again assuming $\varphi \in C^\infty(\Omega)$,

$$\begin{aligned}
\gamma \left(|\nabla \mathbf{u}_n|^2 s_n - |\nabla \mathbf{u}|^2 s, \varphi\right) &= \gamma \left(\left(|\nabla \mathbf{u}_n|^2 - |\nabla \mathbf{u}|^2 \right) s_n, \varphi\right) + \gamma \left(|\nabla \mathbf{u}|^2 \left(s_n - s\right), \varphi\right) \\
&=: \gamma I + \gamma II.
\end{aligned}$$

We compute

$$I \leq C \left\| \nabla \mathbf{u}_n - \nabla \mathbf{u}\right\|_{L^2} \left\| \nabla \mathbf{u}_n + \nabla \mathbf{u}\right\|_{L^2} \left\| s_n\right\|_{L^\infty} \left\| \varphi\right\|_{L^\infty},$$

and

$$II \leq \|\nabla \mathbf{u}\|_{L^2}^2 \, \|s_n - s\|_{L^\infty} \, \|\varphi\|_{L^\infty},$$

which both tend to zero by (2.7.12). In order to identify b in $(2.7.12)_4$, we use Minty's trick. By Lemma 2.7.7(ii),

$$\left(|\Delta s_n|^{p-2} \, \Delta s_n - |\Delta t|^{p-2} \, \Delta t, \Delta s_n - \Delta t\right) \geq 0$$

for all $n \in \mathbb{N}$ and all $t \in W^{2,p}(\Omega)$. Using equation (2.7.10), we get

$$\begin{aligned} -\frac{2\alpha\varepsilon}{\eta_\varepsilon} \left(\nabla s_n, \nabla s_n\right) - \frac{\gamma}{\eta_\varepsilon} \left(|\nabla \mathbf{u}_n|^2 \, s_n, s_n\right) - \frac{\alpha}{2\varepsilon\eta_\varepsilon} \left(s_n - 1, s_n\right) \\ - \left(|\Delta s_n|^{p-2} \, \Delta s_n, \Delta t\right) - \left(|\Delta t|^{p-2} \, \Delta t, \Delta s_n - \Delta t\right) \geq 0 \end{aligned}$$

Using (2.7.12), we can pass to the limit, getting

$$\begin{aligned} -\frac{2\alpha\varepsilon}{\eta_\varepsilon} \left(\nabla s, \nabla s\right) - \frac{\gamma}{\eta_\varepsilon} \left(|\nabla \mathbf{u}|^2 \, s, s\right) - \frac{\alpha}{2\varepsilon\eta_\varepsilon} \left(s - 1, s\right) \\ - \left(b, \Delta t\right) - \left(|\Delta t|^{p-2} \, \Delta t, \Delta s - \Delta t\right) \geq 0, \end{aligned}$$

which, using equation (2.7.8), can be rewritten as

$$\left(b - |\Delta t|^{p-2} \, \Delta t, \Delta s - \Delta t\right) \geq 0$$

for all $t \in W^{2,p}(\Omega)$. Fixing $\varphi \in W^{2,p}(\Omega)$ and setting $t := s - \lambda\varphi$ for $\lambda \geq 0$, we obtain

$$\left(b - |\Delta s - \lambda\Delta\varphi|^{p-2} \, (\Delta s - \lambda\Delta\varphi), \Delta\varphi\right) \geq 0,$$

which, for $\lambda \to 0$, gives

$$\left(b - |\Delta s|^{p-2} \, \Delta s, \Delta\varphi\right) \geq 0,$$

for all $\varphi \in W^{2,p}(\Omega)$. And replacing φ by $-\varphi$, we see that equality holds, so

$$\left(b, \Delta\varphi\right) = \left(|\Delta s|^{p-2} \, \Delta s, \Delta\varphi\right)$$

for all $\varphi \in W^{2,p}(\Omega)$, which identifies the limit. □

2.8 Computational Studies

2.8.1 Implementation of the Penalisation & Splitting Algorithm

To implement Algorithm 2.6.2, we use a simple fixed-point strategy for the Ginzburg-Landau term. So the actual algorithm looks like this,

Algorithm 2.8.1. *Let* $\mathbf{U}^0, \mathbf{G} \in V_h(\Omega, \mathbb{R}^m), S^0 \in V_h(\Omega)$, *and* $0 < \varepsilon, k_\varepsilon, \delta_\varepsilon, \vartheta \ll 1$ *be given. For* $n = 1, \ldots$ *do*

(1) *Set* $\mathbf{U}^{n,0} := \mathbf{U}^{n-1}$, $S^{n,0} := S^{n-1}$

(2) *For* $l = 0, \ldots, \vartheta$ *do*

(a) Compute $\mathbf{U}^{n,l+1} \in V_h(\Omega, \mathbb{R}^m)$ *such that for all* $\mathbf{W} \in V_h(\Omega, \mathbb{R}^m)$ *it holds that*

$$\begin{aligned}\gamma \left(\left(\left(S^{n-1}\right)^2 + k_\varepsilon \right) \nabla \mathbf{U}^{n,l+1}, \nabla \mathbf{W} \right) + \lambda \left(\mathbf{U}^{n,l+1} - \mathbf{G}, \mathbf{W} \right) \\ + \frac{1}{\delta_\varepsilon} \left(\left(\left| \mathbf{U}^{n,l} \right|^2 - 1 \right) \mathbf{U}^{n,l+1}, \mathbf{W} \right) &= 0.\end{aligned}$$

(b) Compute $S^{n,l+1} \in V_h(\Omega)$ *such that for all* $W \in V_h(\Omega)$ *it holds that*

$$\begin{aligned}2\alpha\varepsilon \left(\nabla S^{n,l+1}, \nabla W \right) + \gamma \left(\left| \nabla \mathbf{U}^{n-1} \right|^2 S^{n,l+1}, W \right) \\ + \frac{\alpha}{2\varepsilon} \left(S^{n,l+1} - 1, W \right)_h &= 0.\end{aligned}$$

(3) *Set* $\mathbf{U}^n := \mathbf{U}^{n,l+1}$.

For the subsequent computations, we choose $\vartheta = 3$.

2.8.2 Algorithm for Chromaticity and Brightness

So far, our algorithms can only process the chromaticity component of a given image. To process real images, a suitable scheme for the brightness component has to be implemented. We suggest to amend Ambrosio and Tortorelli's energy to $AT_\varepsilon(\mathbf{u}, v, s) : H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega) \times H^1(\Omega) \to [0, +\infty]$

$$\begin{aligned}AT_\varepsilon(\mathbf{u}, v, s) \;:=\; & \frac{\gamma}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \frac{\lambda}{2} \int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x} \\ & + \frac{\gamma_1}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla v|^2 \mathrm{d}\mathbf{x} + \frac{\lambda_1}{2} \int_\Omega |v - b|^2 \mathrm{d}\mathbf{x} \\ & + \alpha \int_\Omega \varepsilon |\nabla s|^2 + \frac{1}{4\varepsilon}(1-s)^2 \mathrm{d}\mathbf{x},\end{aligned} \tag{2.8.1}$$

with $\gamma, \gamma_1, \alpha, \lambda, \lambda_1$ positive constants and $b, v \in L^\infty(\Omega) \cap H^1(\Omega)$ the brightness component of the original and the processed image, respectively (normalised to lie in $[0, 1]$). So, we add a smoothing and a fidelity term for the brightness component in the second line of (2.8.1). The idea here is that the smoothing term for the chromaticity component forces $|s|$ to be small whenever $|\nabla \mathbf{u}|$ is large, while the smoothing term for the brightness component does the same whenever $|\nabla v|$ is large. So we expect $\{s \approx 0\}$ to approximate the union of the essential jump sets of the chromaticity and the brightness component.

This necessitates the adaptation of the optimisation problem for s as well as the solution of a third optimisation problem, which we place between the two existing ones.

If we process an image with more noise in the chromaticity as in the brightness, as is usually the case with images from digital cameras, we can therefore choose to give more weight to the information of the brightness component, and the chromaticity component will profit from the better information of the brightness component through the joint edge set, as illustrated in Example 2.6.

Remark 2.8.2. *In [73], Kawohl finds a connection between two different models in image processing: the variational Mumford-Shah approach and the anisotropic diffusion*

approach by Perona and Malik [90]. He shows that formally applying several simplifications to the former leads to a version of latter. Below, we apply the same procedures to our new model (2.8.1), which we slightly modify for convenience to give $\widetilde{AT}_\varepsilon(\mathbf{u}, v, s) : H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega) \times H^1(\Omega) \to [0, +\infty]$

$$\begin{aligned}\widetilde{AT}_\varepsilon(\mathbf{u}, v, s) \;:=\; & \gamma \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \lambda \int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x} \\ & + \gamma_1 \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla v|^2 \mathrm{d}\mathbf{x} + \lambda_1 \int_\Omega |v - b|^2 \mathrm{d}\mathbf{x} \\ & + \int_\Omega \varepsilon |\nabla s|^2 + \frac{1}{4\varepsilon}(1-s)^2 \mathrm{d}\mathbf{x},\end{aligned} \tag{2.8.2}$$

Variation of this modified functional $\widetilde{AT}_\varepsilon(\mathbf{u}, v, s)$ *with respect to* s *gives the optimality condition*

$$\gamma s |\nabla \mathbf{u}|^2 + \gamma_1 s |\nabla v|^2 - \varepsilon \Delta s + \frac{1}{4\varepsilon}(s-1) = 0.$$

Upon choosing $\varepsilon = o(h^2)$*, we can neglect the term* $\varepsilon \Delta s$*, since on a discrete level,* $\Delta s = O(h^2)$*. (This is supposed to serve as a motivation only — in light of our* Γ*-convergence analysis, this choice of* ε *makes little sense). This motivates the choice of*

$$s := \left(1 + 4\varepsilon \left(\gamma |\nabla \mathbf{u}|^2 + \gamma_1 |\nabla v|^2\right)\right)^{-1},$$

which can be plugged into (2.8.2)*, leading, upon neglecting* $\int_\Omega \varepsilon |\nabla s|^2$ *as above and setting* $k_\varepsilon := 0$*, to the functional*

$$\int_\Omega \frac{\gamma |\nabla \mathbf{u}|^2 + \gamma_1 |\nabla v|^2}{1 + 4\varepsilon \left(\gamma |\nabla \mathbf{u}|^2 + \gamma_1 |\nabla v|^2\right)} \mathrm{d}\mathbf{x} + \lambda \int_\Omega |\mathbf{u} - \mathbf{g}|^2 \mathrm{d}\mathbf{x} + \lambda_1 \int_\Omega |v - b|^2 \mathrm{d}\mathbf{x},$$

with $|\mathbf{u}| = 1$ *a.e. This leads to the Euler equations*

$$\begin{aligned} \gamma \mathrm{div}\left(\frac{\nabla \mathbf{u}}{\left(1 + 4\varepsilon \left(\gamma |\nabla \mathbf{u}|^2 + \gamma_1 |\nabla v|^2\right)\right)^2}\right) &= \lambda(\mathbf{u} - \mathbf{g}) + \mu \mathbf{u}, \\ |\mathbf{u}|^2 &= 1 \text{ a.e.}, \\ \gamma_1 \mathrm{div}\left(\frac{\nabla v}{\left(1 + 4\varepsilon \left(\gamma |\nabla \mathbf{u}|^2 + \gamma_1 |\nabla v|^2\right)\right)^2}\right) &= \lambda_1 (v - b), \end{aligned}$$

for a Lagrange multiplier $\mu : \Omega \to \mathbb{R}$*. This suggests the following modified Perona-Malik approach for colour images using the CB colour model (in the Perona-Malik context, there is usually no fidelity term):*

$$\begin{aligned} \mathbf{u}_t - \mathrm{div}\left(g_1\left(|\nabla \mathbf{u}|^2, |\nabla v|^2\right) \nabla \mathbf{u}\right) &= \mu \mathbf{u}, \\ |\mathbf{u}| &= 1 \text{ a.e.}, \\ v_t - \mathrm{div}\left(g_2\left(|\nabla \mathbf{u}|^2, |\nabla v|^2\right) \nabla v\right) &= 0, \end{aligned} \tag{2.8.3}$$

with no-flux boundary conditions and g_i *positive and tending to zero for* $|\nabla \mathbf{u}|, |\nabla v| \to +\infty$*. The idea is, of course, to have very little diffusion at edges and much more elsewhere. According to the above formal derivation, the natural choice for* g_i *would be*

$g_i\left(|\nabla \mathbf{u}|^2, |\nabla v|^2\right) := \left(1 + \gamma |\nabla \mathbf{u}|^2 + \gamma_1 |\nabla v|^2\right)^{-2}$ *for* $i \in \{1, 2\}$. *Like with the Mumford-Shah functional, this means that chromaticity and brightness share a common edge set. In fact, for either* $|\nabla \mathbf{u}|$ *or* $|\nabla v|$ *large enough, backward diffusion has to be expected, which steepens profiles (enhances edges), according to results on the standard Perona-Malik approach, assuming classical solutions.*

Using Graßmann's identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$ *for* $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^m$, $(2.8.3)_1$ *and* $(2.8.3)_2$ *can be recast into*

$$\mathbf{u}_t - \mathbf{u} \times (\mathbf{u} \times \mathbf{DG}(\mathbf{u})) = 0,$$

with $(\mathbf{DG}(\mathbf{u}), \boldsymbol{\varphi}) := \int_\Omega g_1\left(|\nabla \mathbf{u}|^2, |\nabla v|^2\right) \nabla \mathbf{u} \cdot \nabla \boldsymbol{\varphi} \mathrm{d}\mathbf{x}$ *for all* $\boldsymbol{\varphi} \in H^1(\Omega, \mathbb{R}^m)$, *a type of problem which was numerically treated in [13].*

2.8.3 Academic Images, Splitting & Projection

All arrows below are scaled in length to fit the plots. Since Algorithm 2.4.1 ensures the sphere constraint exactly at nodal points, any appearance of changing length is due to perspective. Note that what we call h below is in fact the length of the two shorter sides of the rectangular triangles in our triangulations; i.e., it is shorter than the actual diameter of the triangles (by a factor of $\sqrt{2}$).

Example 2.1. *Let* $\Omega := (0, 1)^2$ *and* $\mathbf{G}$ *as in the left plot in Figure 2.3. The right picture shows a section along* $x = 0.5$, *where the z-values of the two regions are the closest. We use a triangulation consisting of* 2^{2*8} *halved squares (along the direction* $(1, 1)$*); i.e.,* 131072 *triangles, with* 66049 *nodes, and* $h = 2^{-8} \approx 4 * 10^{-3}$. *The initial values for* $\mathbf{U}$ *and* S *are* $\mathbf{U}_0 \equiv \mathbf{G}$ *and* $S_0 \equiv 0.5$, *respectively. We choose* $\gamma = 1.2$, $\alpha = 0.5$, $\lambda = 2 * 10^3$, $\varepsilon = 6 * 10^{-4}$, *and* $k_\varepsilon = 10^{-6}$ *(parameters chosen by experiment).*

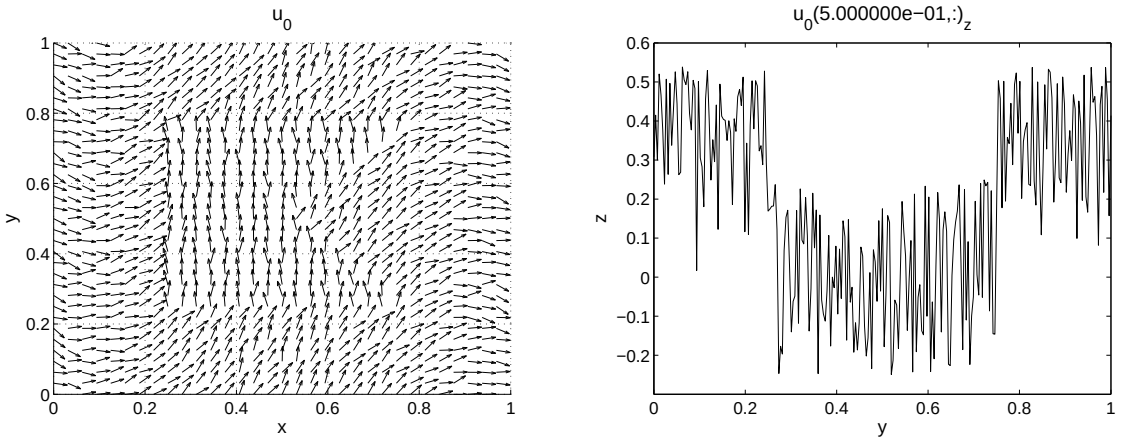

Figure 2.3: Example 2.1: Original image (left) and z-values of a vertical section through it ($x = 0.5$, right).

Figure 2.3 shows the initial values, Figure 2.4 the result after 10 iterations of our proposed algorithm. Figure 2.5 shows the detected edge set and Figure 2.6 the Ambrosio-Tortorelli energy over time.

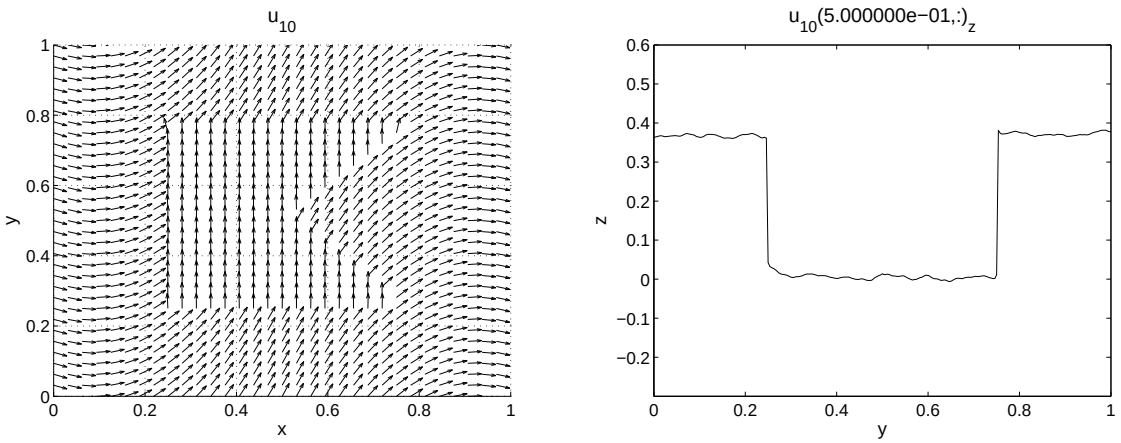

Figure 2.4: Example 2.1: Image and section after 10 iterations.

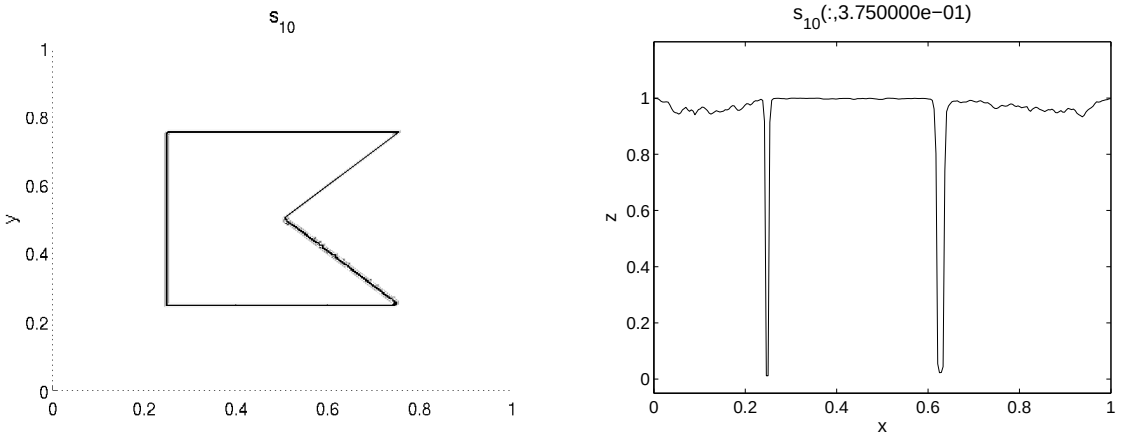

Figure 2.5: Example 2.1: Edge set (left) and horizontal section through it ($y = 0.375$, right) after 10 iterations.

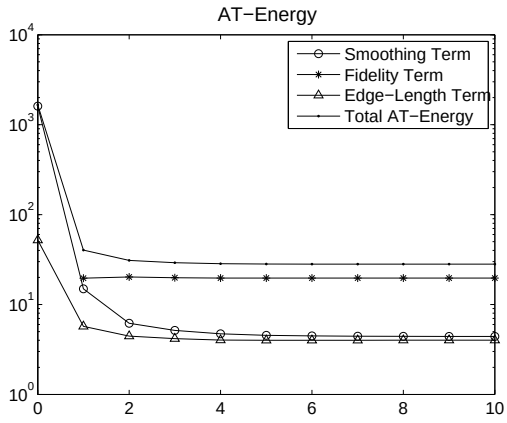

Figure 2.6: Example 2.1: Ambrosio-Tortorelli Energy (10 iterations, logarithmic plot).

The next example numerically studies simultaneous blowup behaviour for the $W^{1,\infty}$-norm of iterates $\{\mathbf{U}_n, S_n\}$.

Example 2.2. *Let Ω, S_0, and $\mathbf{G}$ be as above. However, $\mathbf{U}_0$ is chosen to be $\mathbf{U}_0 \equiv (0,0,1)$. We first use a triangulation consisting of 2^{2*r}, $r = 8$ halved squares as above, and later use coarser ones ($r \in \{5,\dots,8\}$) for comparison. Parameters are as above, except for $\varepsilon = h/6 = 2^{-r}/6$.*

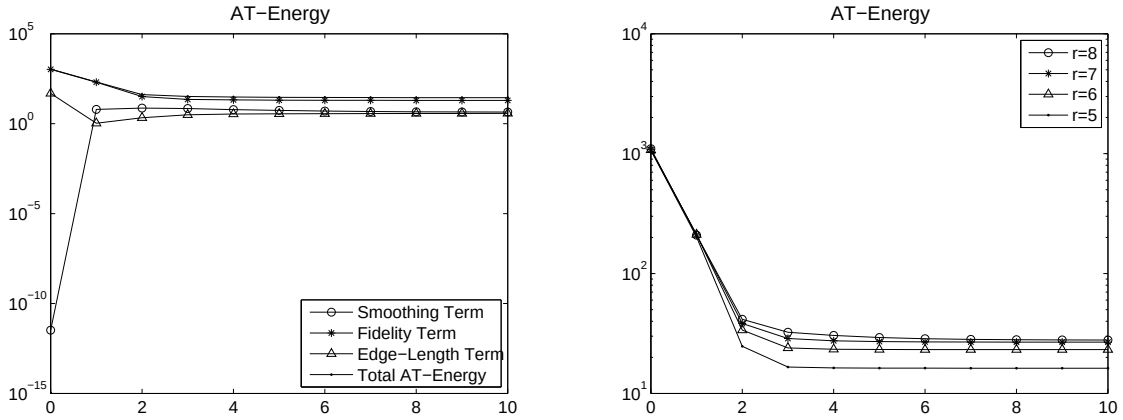

Figure 2.7: Example 2.2: Ambrosio-Tortorelli energy, 10 steps, for $r = 8$ (left) and $r \in \{5,\dots,8\}$ (right), y-logarithmic plots.

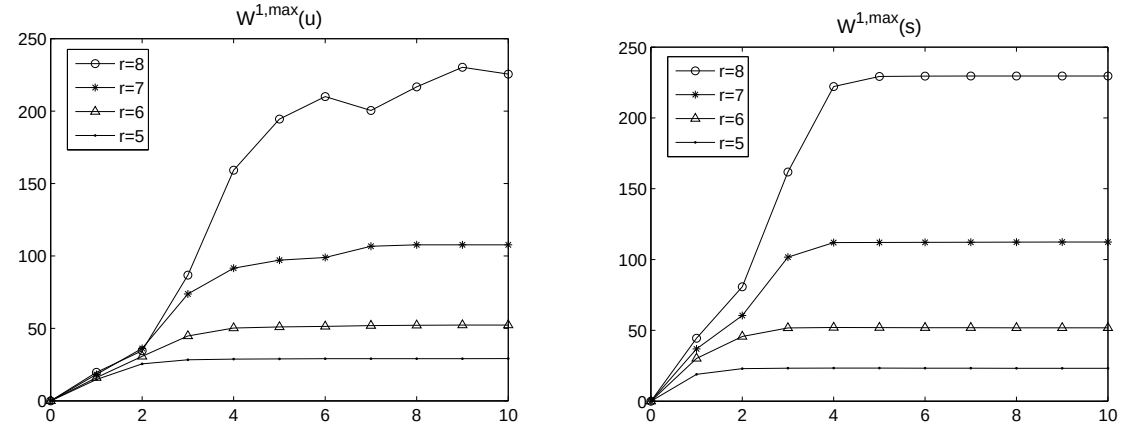

Figure 2.8: Example 2.2: $W^{1,\infty}$-norm of $\mathbf{U}$ and S for $r \in \{5,\dots,8\}$.

Results for $r = 8$ turn out to be almost identical to those in the first example. Figure 2.7 shows the total Ambrosio-Tortorelli energy for $r \in \{5,\dots,8\}$ (which are very close for all r). Finally, Figure 2.8 shows the $W^{1,\infty}$-norm of $\mathbf{U}$ and S for $r \in \{5,\dots,8\}$.

We experience blowup behaviour for $\{\mathbf{U}_n\}$ and $\{S_n\}$, apparently driven by the fidelity term; and $\{S_n\}$ appears to be one iteration ahead of $\{\mathbf{U}_n\}$ in terms of blowup behaviour.

The next example numerically studies blowup behaviour for the $W^{1,\infty}$-norm of iterates $\{\mathbf{U}_n, S_n\}$ in the absence of a fidelity term; i.e., $\lambda = 0$. This is motivated by blowup results

for harmonic maps (to the sphere), see e.g. [95, 97, 98, 99, 68, 12]. In particular, it is known that for $d = 2$, singularities only appear for large initial energy. And any harmonic map (for general d) is smooth outside a set whose $(d-2)$-dimensional Hausdorff measure is zero, see [94, 95, 69, 58, 18, 77, 76].

Example 2.3. *We use the same domain and triangulation as in Example 2.2 with* $r \in \{5, \dots, 8\}$. *Let* $\gamma = 1 = \alpha$, $\lambda = 0$, $\varepsilon = h/6$, *and* $k_\varepsilon = 10^{-6}$. *We use two sets of initial data for* $\mathbf{U}$ *and* S, *which are shown in Figures 2.9 and 2.12 (leftmost column). In both cases,* $\mathbf{U}_0$ *is constantly* $(0, 0, 1)$ *in the periphery of the image,* $(0, 0, -1)$ *at the centre, and varying continuously inside a circle around the centre. In the first case, we choose* $S = 0$ *at the centre,* $S = 1$ *in the periphery, and smoothly varying in between; in the second case, we choose* $S = 1$ *at the centre,* $S = 0$ *in the periphery, and smoothly varying in between.*

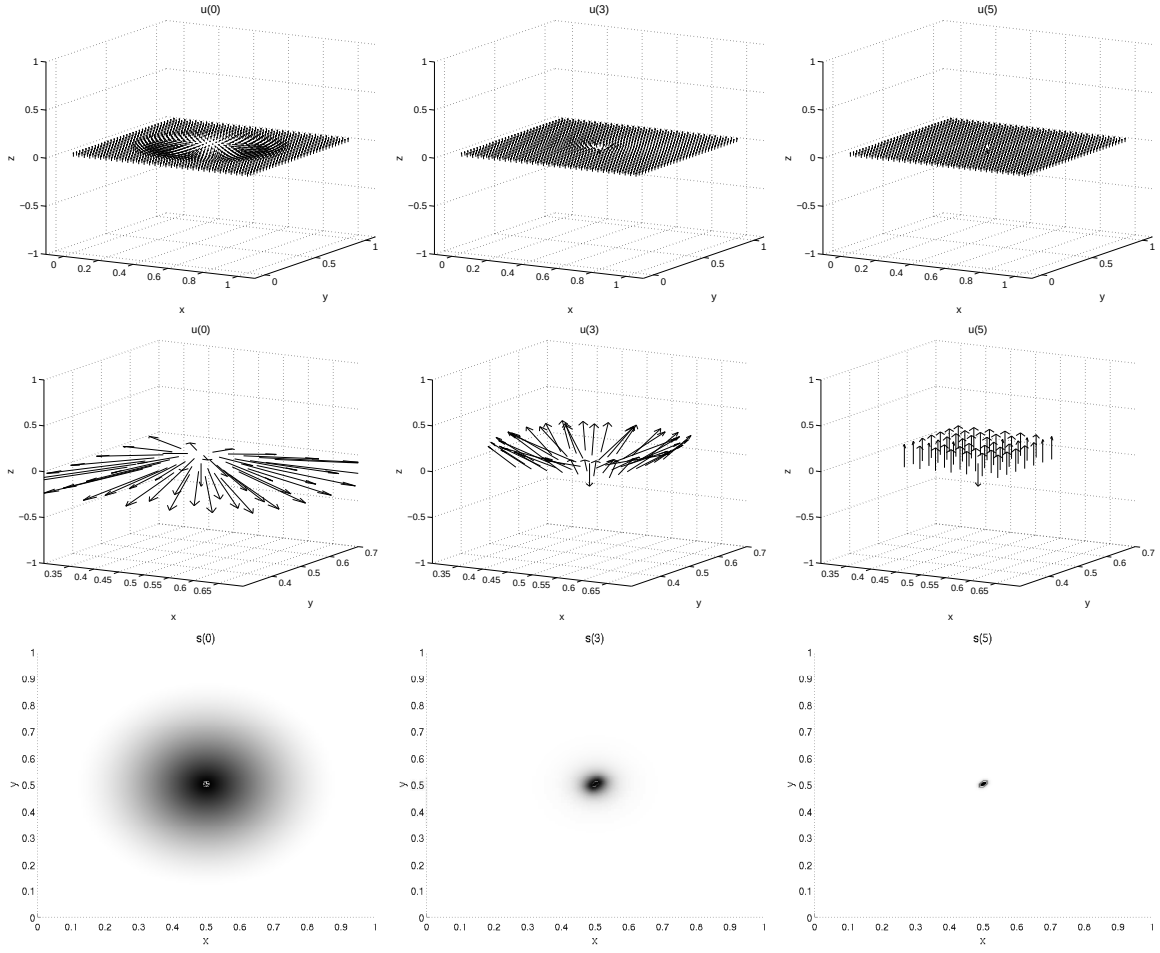

Figure 2.9: Example 2.3: $\mathbf{U}_n$ (top), detail of $\mathbf{U}_n$ (middle), and full image of S_n (bottom) for $n \in \{0, 3, 5\}$.

Figure 2.9 shows iterates $n \in \{0, 3, 5\}$ for $r = 8$, and crops of iterates $\mathbf{U}_n$, Figure 2.10 shows the total energy for $r \in \{5, \dots, 8\}$, while Figure 2.11 shows the $W^{1,\infty}$-norms of $\mathbf{U}_n$ and S_n for $r \in \{5, \dots, 8\}$, which both show blowup behaviour. This time it is $\mathbf{U}_n$ which

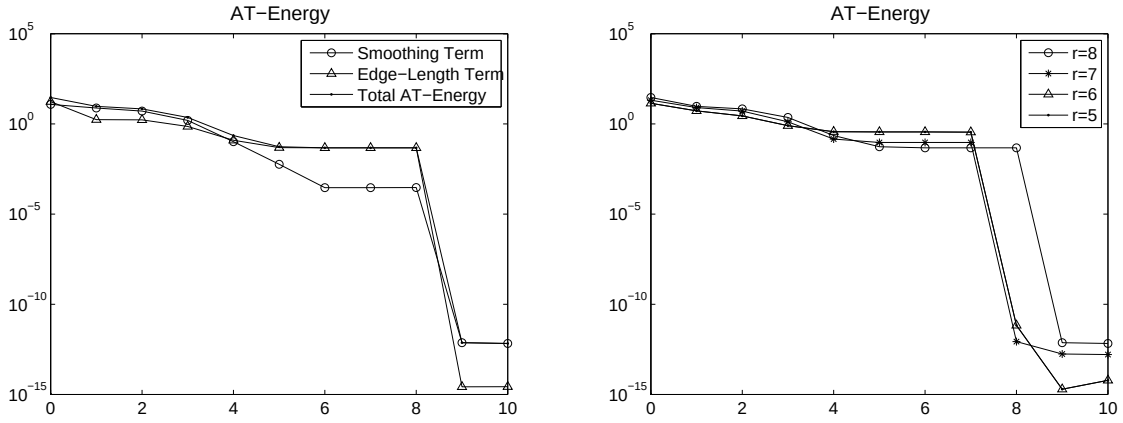

Figure 2.10: Example 2.3: Ambrosio-Tortorelli energy, 10 steps, for $r = 8$ (left) and $r \in \{5, \dots, 8\}$ (right), y-logarithmic plots.

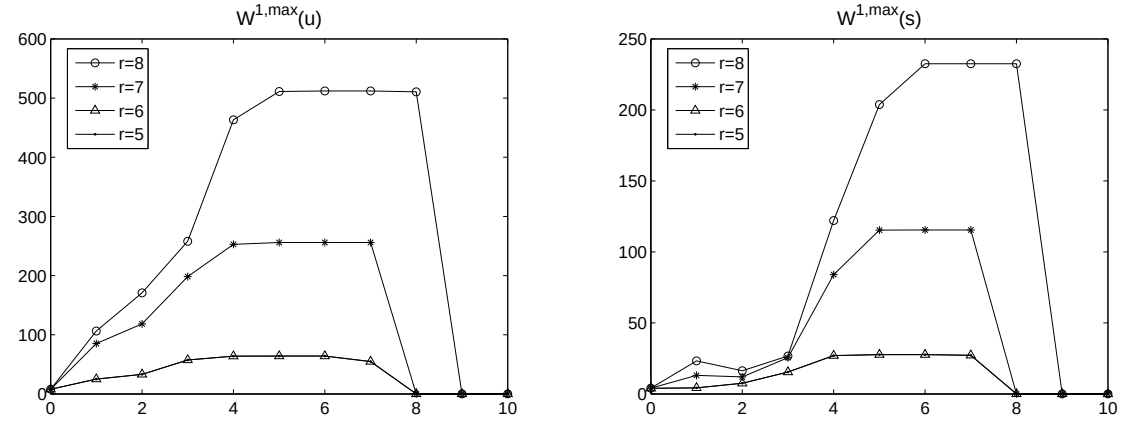

Figure 2.11: Example 2.3: $W^{1,\infty}$-norm of $\mathbf{U}$ and S for $r \in \{5, \dots, 8\}$.

appears one step ahead of S_n with respect to blowup behaviour. Depending somewhat on r, the system matrices become close to singular after 6–7 iterations, so after this point, the results can no longer be expected to be reliable. The arrow at the centre of $\mathbf{U}$ at this point still points down, while the rest of $\mathbf{U}$ points up. The variable S, on the other hand, becomes 1 everywhere, except for the centre, where it stays 0. After breakdown, the arrows move erratically, but perfectly synchronised with one another.

The next example uses the same setting and the same initial data for $\mathbf{U}$, but avoids blowup behaviour through a different choice of initial data for S.

Example 2.4. *Except for the initial data for S we use exactly the same setting as in Example 2.3. This time, we choose $S = 1$ at the centre, $S = 0$ in the periphery, and smoothly varying in between.*

Figure 2.12 shows iterates $n \in \{0, 3, 6\}$ for $r = 8$, and crops of iterates $\mathbf{U}_n$, Figure 2.13 shows the total energy for $r \in \{5, \dots, 8\}$, while Figure 2.14 shows the $W^{1,\infty}$-norms of $\mathbf{U}_n$ and S_n for $r \in \{5, \dots, 8\}$, which this time stay finite. The arrows of $\mathbf{U}$ all point

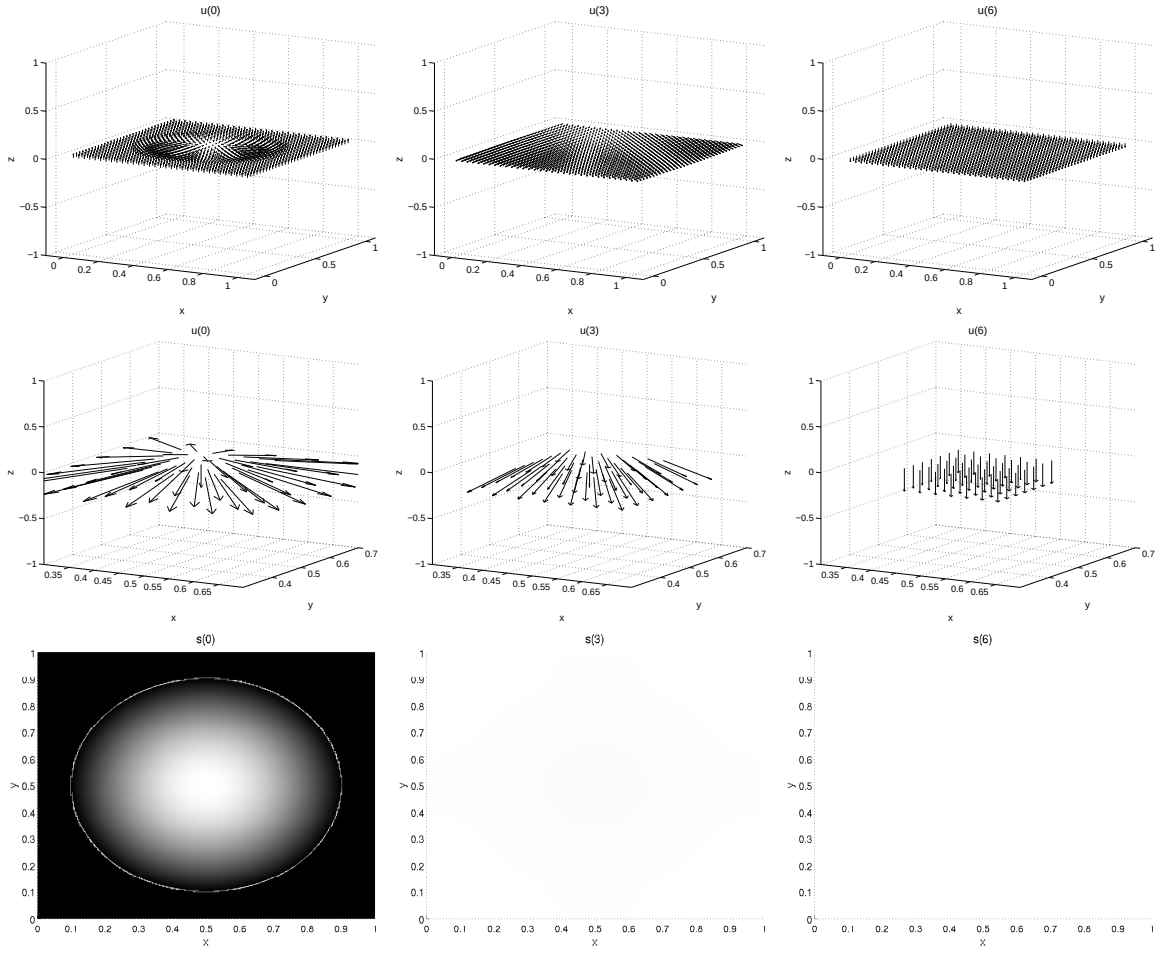

Figure 2.12: Example 2.3: $\mathbf{U}_n$ (top), detail of $\mathbf{U}_n$ (middle), and S_n (bottom) for $n \in \{0, 3, 6\}$.

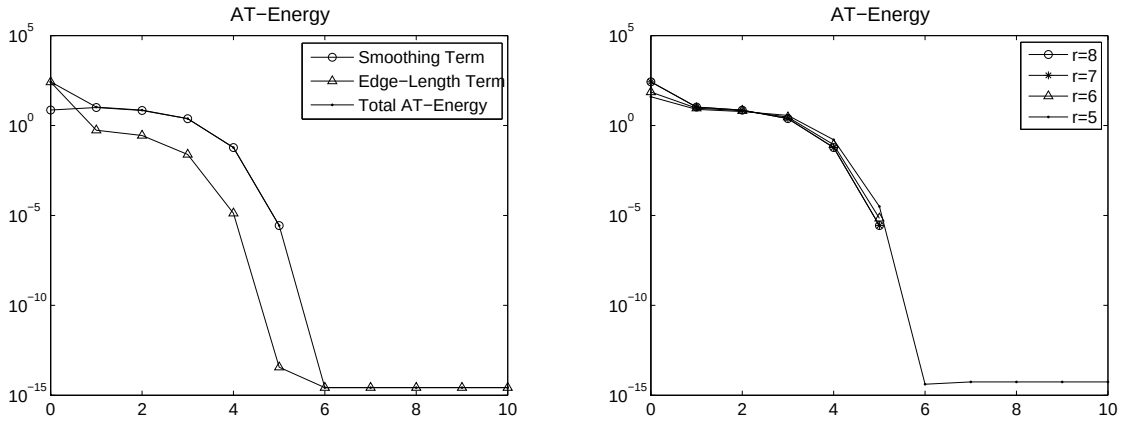

Figure 2.13: Example 2.3: Ambrosio-Tortorelli energy, 10 steps, for $r = 8$ (left) and $r \in \{5, \ldots, 8\}$ (right), y-logarithmic plots.

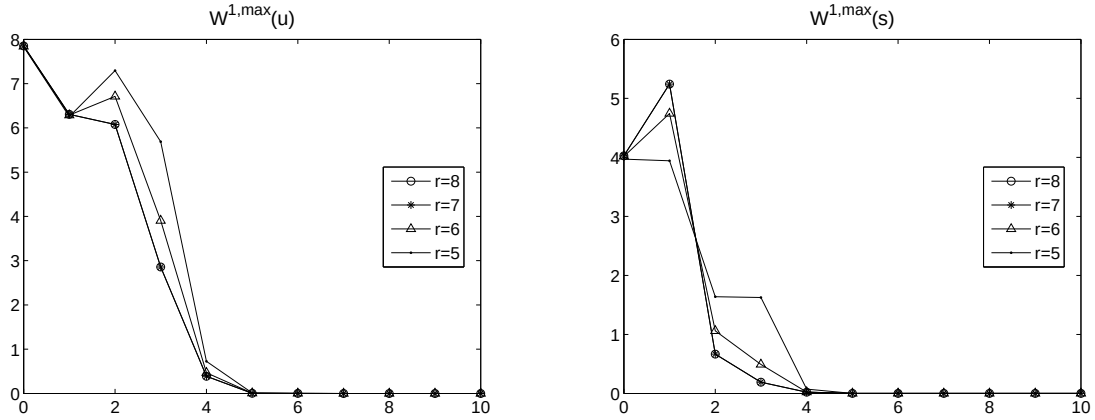

Figure 2.14: Example 2.3: $W^{1,\infty}$-norm of $\mathbf{U}$ and S for $r \in \{5, \ldots, 8\}$.

down at the end, while S is 1 everywhere. After 6 iterations, the system matrices again become close to singular; in this case, however, iterates do not change dramatically after this point, if at all.

2.8.4 Academic Images, Penalisation & Splitting

The next example studies the same setting as Example 2.1, this time with Algorithm 2.6.2; i.e., the sphere constraint is enforced by penalisation instead of projection. Again, all arrows are scaled in length to fit the plots.

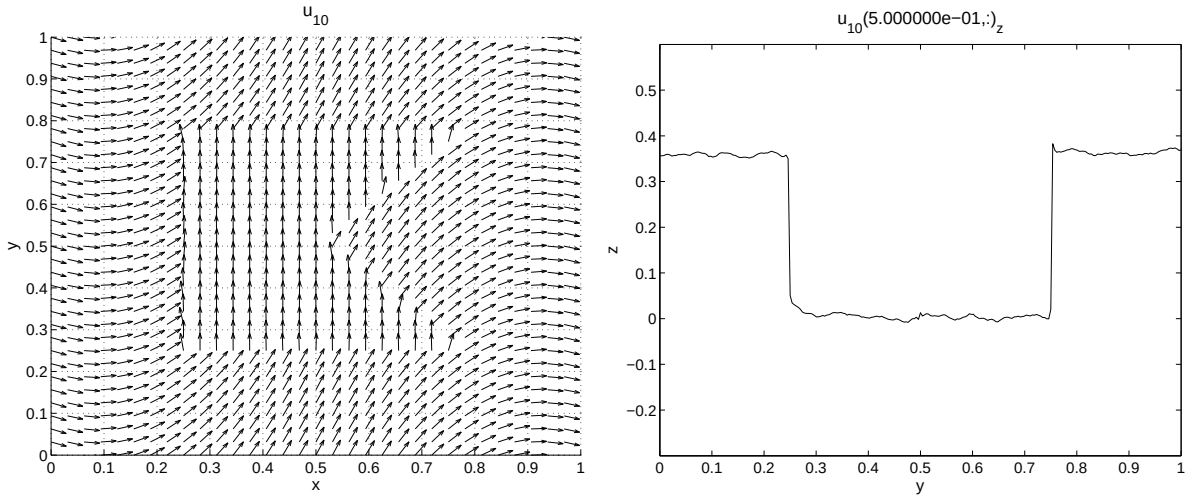

Figure 2.15: Example 2.1: Image and section after 10 iterations.

Example 2.5. *The setting is as in Example 2.1. Parameters are* $\gamma = 1.2$, $\alpha = 0.5$, $\lambda = 2 * 10^3$, $\varepsilon = 10^{-3}$, $k_\varepsilon = 10^{-6}$, *and* $\delta_\varepsilon = 0.1$ *(chosen by experiment).*

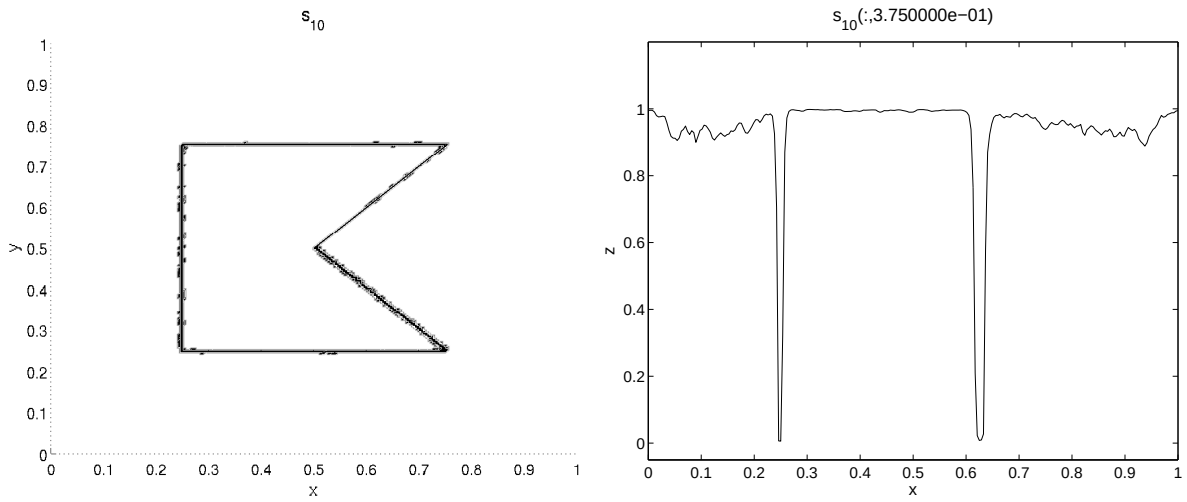

Figure 2.16: Example 2.5: Edge set (left) and horizontal section through it ($y = 0.375$, right) after 10 iterations.

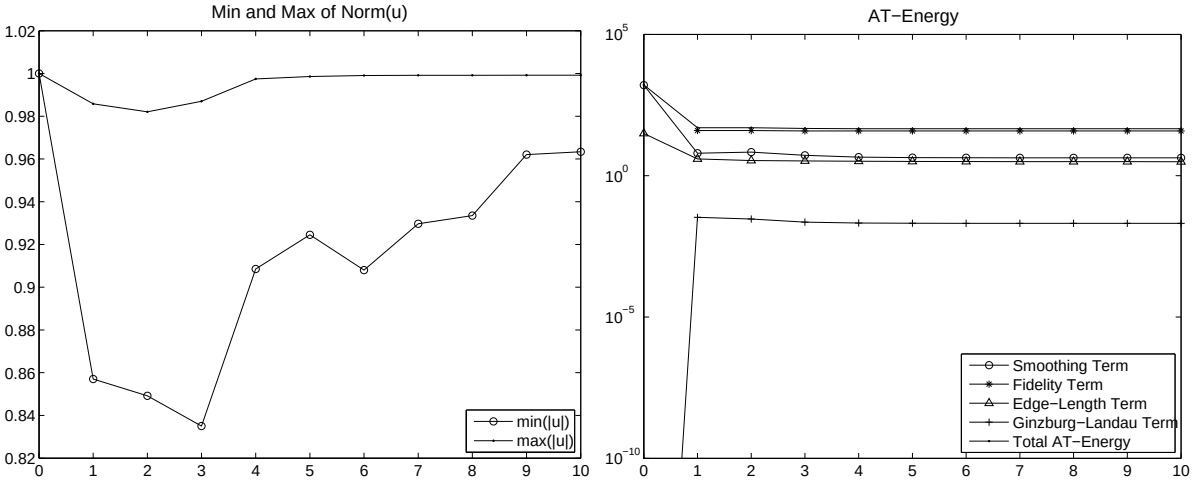

Figure 2.17: Example 2.5: min and max of $|\mathbf{U}|$ (left), and Ambrosio-Tortorelli Energy (right) for 10 iterations.

Figure 2.15 shows the result after 10 iterations, Figure 2.16 shows the detected edge set after 10 iterations, while Figure 2.17 shows the global minimum and maximum of $|\mathbf{U}|$ and the Ambrosio-Tortorelli energy over time.

For δ_ε between about $5*10^{-3}$ and at least 10^2, the results are qualitatively very similar to the ones in Example 2.1, but the detected edge set is less exact, and $|\mathbf{U}|$ can be quite a bit shorter than 1. For δ_ε smaller than $5 * 10^{-3}$ (which would be advantageous for the accuracy of $|\mathbf{U}|$), the results break down, which is in accordance with our theoretical results.

2.8.5 Real Image, Splitting & Projection

Example 2.6. *We try our algorithm on a small photograph (*399×299 *pixels), as shown in Figure 2.18. We choose* $\Omega := (0, 399/299) \times (0, 1)$*, whence* $h = 1/298 \approx 3 * 10^{-3}$*, the pixels are used as nodes, each square of* 4 *pixels giving rise to two triangles. We further choose* $S_0 \equiv 1$ *and add two different kinds of noise to the image:*

(1) *RGB noise:* $R = R_0 + 0.3 * \text{randn}$*, and* G *and* B *analogously, where* randn *are pseudo-random values drawn from the standard normal distribution. After this operation, we crop* R*,* G*, and* B *to lie in* $[0, 1]$ *(where* R_0, G_0, B_0 *were scaled to lie). This is shown in Figure 2.18.*

(2) *CB noise, mainly in the chromaticity component:* $\mathbf{C} = \mathbf{C}_0 + 0.5 * \text{randn} * \mathbf{C}_0 \times [1, 1, 1] \in \mathbb{S}^2$*, and* $B = B_0 + 0.01 * \text{randn}$*. After this operation,* $\mathbf{C}$ *is projected onto the sphere, and* B *is cropped to lie in* $[0, 1]$*. This is shown in Figure 2.22.*

Our CB algorithm was in both cases compared to a channelwise RGB computation for the same image, with all channels sharing the same edge set. Parameters were chosen as follows (by experiment):

(1) *RGB computation:* $\alpha = 0.3$*,* $\beta = 10^{-2}$*,* $\gamma = 10^3$*,* $\varepsilon = 10^{-4}$*, and* $k_\varepsilon = 10^{-7}$*.*
CB computation: $\alpha = \alpha_1 = 0.5$*,* $\beta = 8 * 10^{-3}$*,* $\gamma = \gamma_1 = 10^3$*,* $\varepsilon = 10^{-4}$*, and* $k_\varepsilon = 10^{-7}$*.*

(2) *RGB computation:* $\alpha = 0.5$*,* $\beta = 5 * 10^{-3}$*,* $\gamma = 50$*,* $\varepsilon = 10^{-4}$*, and* $k_\varepsilon = 10^{-7}$*.*
CB computation: $\alpha = \alpha_1 = 0.3$*,* $\beta = 5 * 10^{-2}$*,* $\gamma = 10^2$*,* $\gamma_1 = 5 * 10^5$*,* $\varepsilon = 10^{-4}$*, and* $k_\varepsilon = 10^{-7}$*.*

Figure 2.18: Example 2.6.1: Original image (left) and image with RGB noise (right).

First, let us look at the computations with RGB noise: Figure 2.18 shows the noisy initial image, and Figure 2.19 the result after 10 iterations. Figure 2.20 shows the detected

Figure 2.19: Example 2.6.1: Image after 10 iterations, RGB (left) and CB (right).

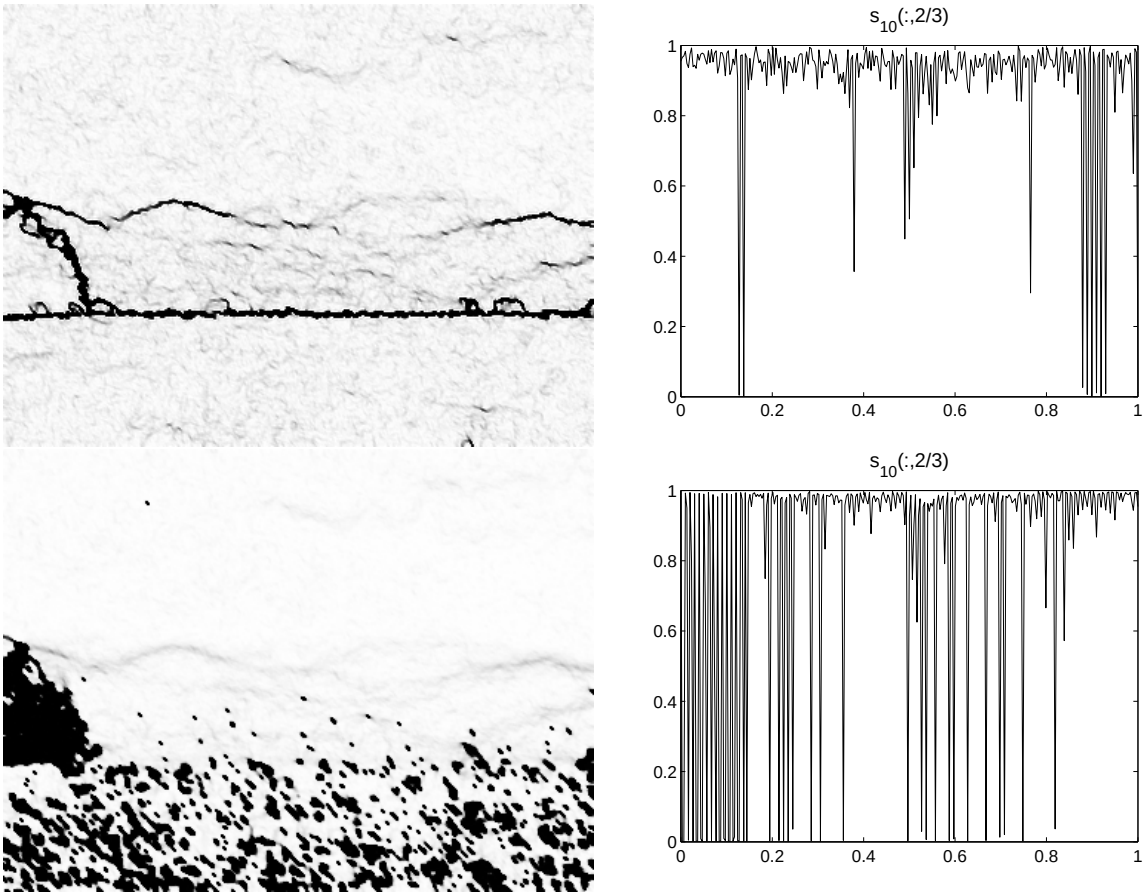

Figure 2.20: Example 2.6.1: Edge set; top: RGB, full image (left) and section (right); bottom: CB, full image (left) and section (right).

edge set, as well as a section through it, and Figure 2.21 the expanded Ambrosio-Tortorelli energy over time. The energy terms labelled "...C" belong to the chromaticity compon-

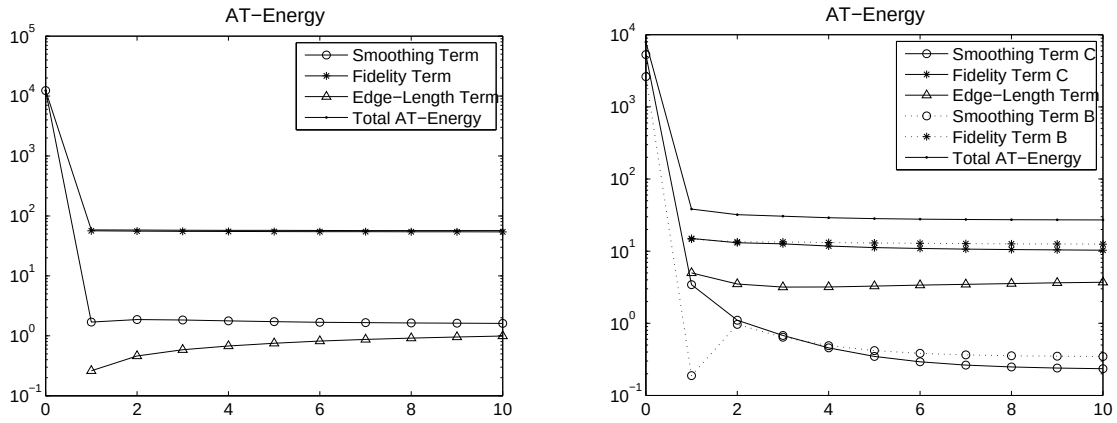

Figure 2.21: Example 2.6.1: Expanded Ambrosio-Tortorelli Energy (10 iterations, y-logarithmic plots), RGB (left) and CB (right).

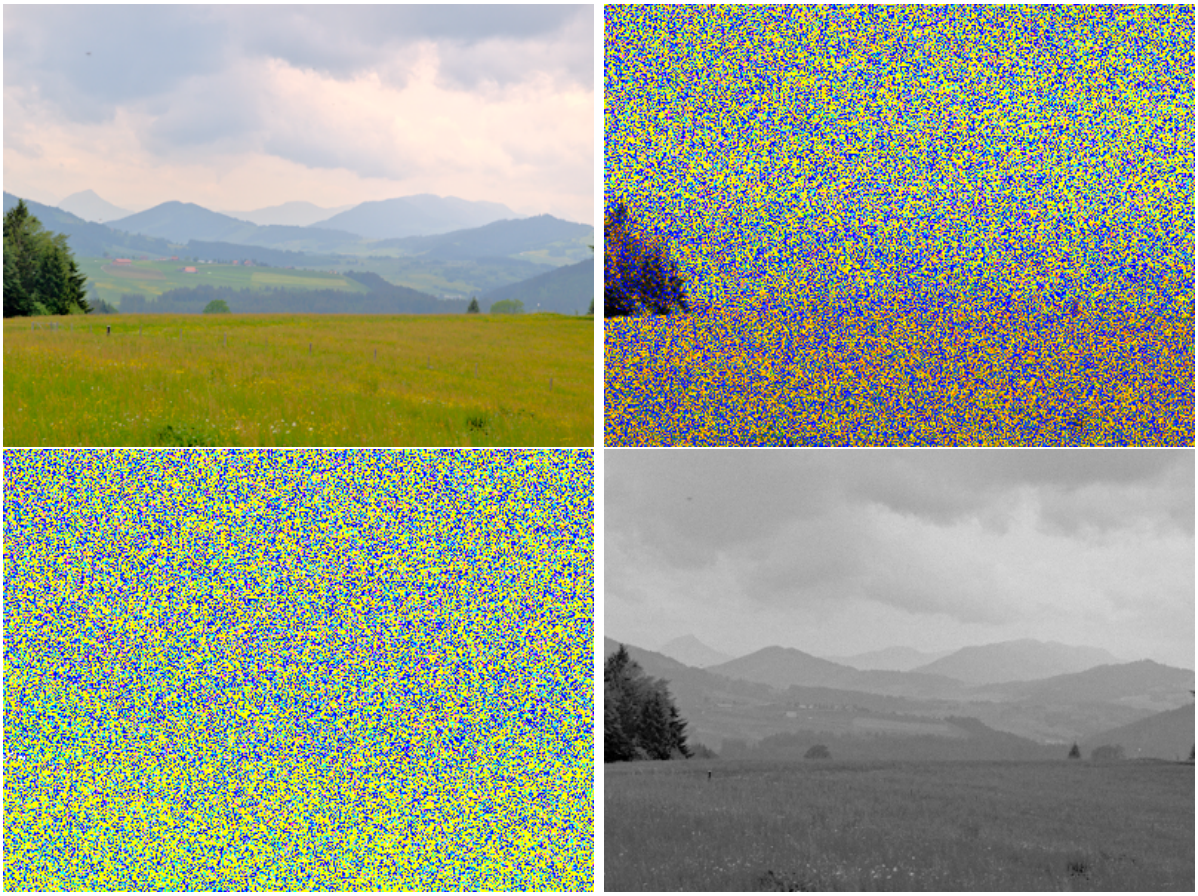

Figure 2.22: Example 2.6.2: Original image and image with CB noise (top), as well as noisy chromaticity (bottom left) and brightness (bottom right) components.

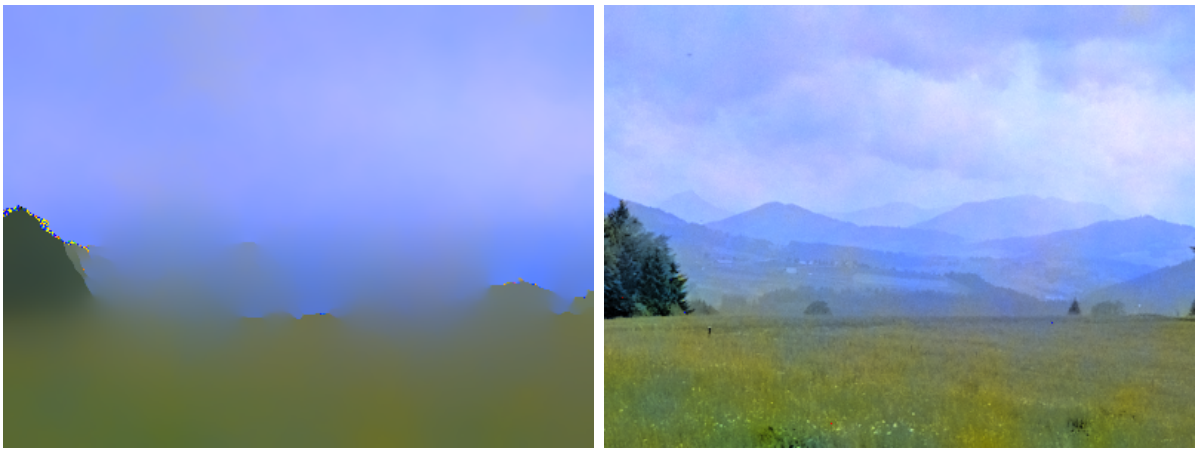

Figure 2.23: Example 2.6.2: Image after 10 iterations, RGB (left) and CB (right).

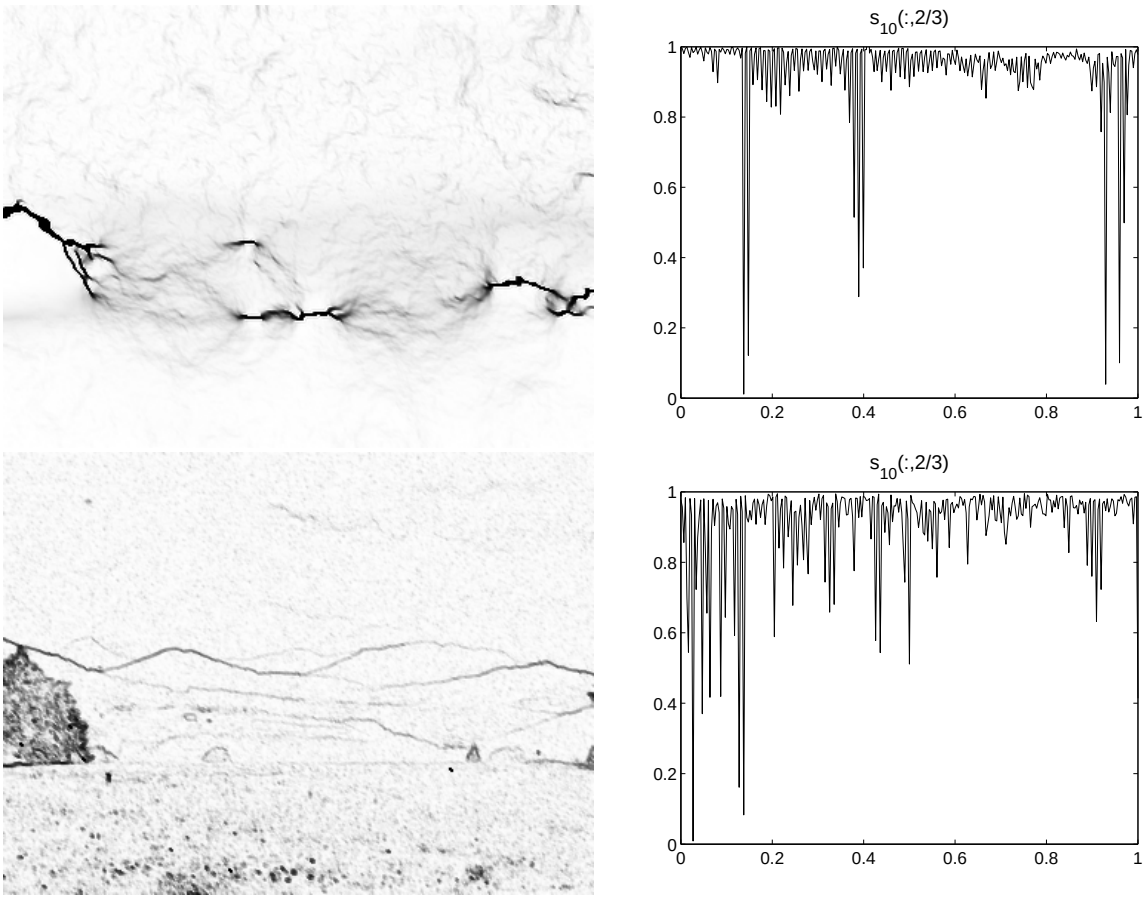

Figure 2.24: Example 2.6.2: Edge set; top: RGB, full image (left) and section (right); bottom: CB, full image (left) and section (right).

ent, those labelled "...B" to the brightness. The channelwise RGB algorithm has the advantage here.

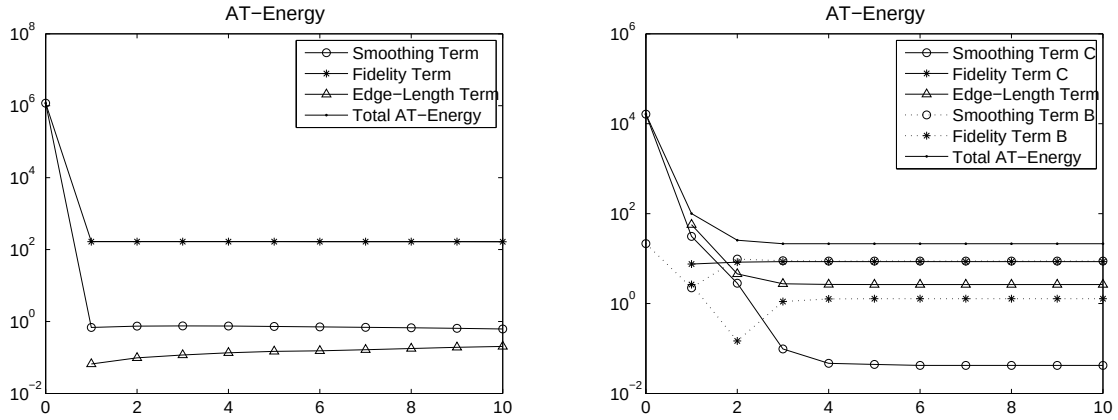

Figure 2.25: Example 2.6.2: Expanded Ambrosio-Tortorelli Energy (10 iterations, y-logarithmic plots), RGB (left) and CB (right).

Next, let us look at the image with CB noise: Figure 2.22 shows the noisy initial image, and Figure 2.23 the result after 10 iterations. Figure 2.24 shows the detected edge set, as well as a section through it, and Figure 2.25 the expanded Ambrosio-Tortorelli energy over time. The CB algorithm has a very clear advantage here.

How realistic any of these cases are is an interesting question that is beyond the scope of this thesis.

2.9 Conclusion

In this chapter, we proposed two splitting strategies based on first-order finite elements, for minimising the Mumford-Shah functional for unit vector fields $\mathbf{u} : \Omega \subset \mathbb{R}^d \to \mathbb{S}^{m-1}$. Both strategies use elliptic approximations to the original functional, based on the Ambrosio-Tortorelli functional.

The first strategy uses a projection idea that enforces the sphere constraint exactly on nodal points, for every iterate $\mathbf{U}_n$. For acute triangulations, every step of the resulting iteration is energy-decreasing, and the algorithm converges weakly (up to subsequences) in $H^1 \times H^1$ to a tuple $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{S}^{m-1}) \times H^1(\Omega)$. For $d = 2$ we can show that s and iterates S_n fulfil $S_n, s \in [-1, 1]$. However, we cannot show that $(\mathbf{u}, s)$ is a stationary point of the Ambrosio-Tortorelli energy for unit vector fields.

The second strategy uses a penalisation approach, adding a Ginzburg-Landau term to the Ambrosio-Tortorelli functional; so iterates $\mathbf{U}_n$ only approximate the sphere constraint. The resulting algorithm converges weakly (up to subsequences) in $H^1 \times H^1$ to a tuple $(\mathbf{u}, s) \in H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega)$, without any mesh-constraint. For $d = 2$ we can also show that $S_n, s \in [-1, 1]$. This allows to get strong convergence (up to subsequences) of iterates $\mathbf{U}_n$ in H^1, which in turn allows to pass to the limit and show that $(\mathbf{u}, s)$ is a stationary point of the Ambrosio-Tortorelli-Ginzburg-Landau energy, and that $s \geq 0$. However, we now have to solve a nonlinear equation in every iteration.

Acknowledgements

For this chapter, we thank Sören Bartels (U Bonn) for his help with the coding, and Giovanni Bellettini (U Roma), Ludwig Gauckler (U Tübingen), Christoph Ortner (U Oxford), Reiner Schätzle (U Tübingen), and Markus Schmuck (MIT Cambridge, MA) for helpful discussions.

Chapter 3

The Mumford-Shah-Euler Functional

In this chapter, we propose, analyse and compare two fully discrete finite element based algorithms for the L^2 gradient flow of the Mumford-Shah-Euler functional for unit vector fields. A special interest in this setup is the curvature term that was added compared to the standard Mumford-Shah functional. The first scheme uses a penalisation strategy, the second uses a Lagrange multiplier, to approximate and enforce the sphere constraint, respectively.

Both schemes are then applied to colour image inpainting in the chromaticity and brightness colour model, and also compared to inpainting with the standard Mumford-Shah functional and channelwise RGB inpainting with the Mumford-Shah-Euler functional.

We again observe that the strategy that only approximates the sphere constraint allows for a better convergence result.

3.1 Introduction

Our interest in this chapter lies in numerics of an energy functional involving a curvature term. One motivation for this choice of topic is analytical interest in studying effects of corresponding flows acting on nontrivial initial data. Another motivation is numerical image processing, in particular image inpainting. The term *image inpainting* was introduced into digital image processing in [17]; it is an artistic synonym for image interpolation and stems from traditional image restoration. In [57], it was demonstrated that for large-scale inpainting problems (i.e., problems with large areas of missing data), it is necessary to introduce curvature information in order to faithfully reconstruct images. One way of doing this is using Euler's elastica curvature model, which was first introduced into computer vision in [85], see also [19]. It was studied for inpainting in [83, 32, 57], see also the surveys [35, 33].

3.1.1 Colour Models

Colour images are commonly represented by functions $\mathbf{u} : \Omega \to \mathbb{R}^3$, where the three dimensions correspond to the primary colours *red*, *green*, and *blue*. However, in many applications it is desirable to treat colour and brightness information separately, e.g.

because of the way images are often recorded (Bayer sensors), and because the human eye treats colour and brightness through different receptors, c.f. Chapter 1. A mathematically appealing way to achieve this separation is the *Chromaticity and Brightness (CB)* colour model, where the chromaticity (colour information) is represented by a function $\mathbf{c} := \mathbf{u}/|\mathbf{u}| : \Omega \to \mathbb{S}^2$, while the brightness is represented by $b := |\mathbf{u}| : \Omega \to \mathbb{R}$, usually scaled to lie in $[0, 1]$.

In [31], it is demonstrated, that this model is advantageous to RGB and to the straightforward HSV (Hue, Saturation, Value) model in colour image denoising and enhancement. In [67] and Chapter 2, we use this model for colour image segmentation and denoising using the Mumford-Shah functional for sphere-valued functions, and again show its advantages in the presence of certain types of colour noise. Other sources include [96, 89, 100, 30, 101, 74, 88, 28, 13, 72] and references therein.

3.1.2 The Mumford-Shah-Euler Functional

Let $R \subset \mathbb{R}^d$ be a polyhedral Lipschitz domain. Given an image $\mathbf{g} : R \to \mathbb{R}^m$, let $K \subset R$ be the inpainting domain (the part of the image where information is missing), $B \subset R \setminus K$ a narrow band around it (whose information will be propagated into the inpainting domain), and $\Omega := K \cup B$.

The *Mumford-Shah-Euler* model is based on minimising a functional proposed in [57], based on ideas from the seminal book [87], see also [85]. A curvature term (Euler term) is added to the (second order) *Mumford-Shah* functional, that was originally proposed in [86] for greyscale image segmentation.

The *Mumford-Shah-Euler* functional for unit vector fields is

$$E(\mathbf{u}, \Gamma) := \frac{\gamma}{2} \int_{\Omega \setminus \Gamma} |\nabla \mathbf{u}|^2 \mathrm{d}\mathbf{x} + \frac{\lambda}{2} \int_B |\mathbf{u} - \mathbf{u}_0|^2 \, \mathrm{d}\mathbf{x} + \sigma \int_\Gamma \left(\alpha + \beta |H|^2\right) \mathrm{d}\mathcal{H}^{d-1}(\mathbf{x}) \,,$$

with H denoting the (mean) curvature of Γ, $\mathbf{u}_0 \in L^\infty(\Omega, \mathbb{S}^{m-1})$, $\psi(t) := (1 - t^2)^2$, and $\sigma := \int_{-1}^{1} \sqrt{2\psi(t)} \mathrm{d}t$. Note that $\psi'(t) = -4t\,(1 - t^2)$, and $\psi''(t) = 12t^2 - 4$. This functional is to be minimised for all $\Gamma \subset \Omega$ closed and C^2, and $\mathbf{u} \in H^1(\Omega \setminus \Gamma, \mathbb{S}^{m-1})$. In this energy, the first term (smoothing term) penalises high gradients outside the *edge set* Γ, the second term (fidelity term) penalises deviations from the original unit vector field $\mathbf{u}_0$ inside the band B, and the last term (Euler term) penalises length and curvature of Γ. The Euler term corresponds to the *Willmore energy*, a generalisation of which was also used in the context of image denoising and segmentation in [50].

Note that given the non-convex nature of the functional above (even without the sphere constraint), uniqueness of minimisers cannot be expected. Indeed, unlike in the context of image segmentation, where a global fidelity term confines minimisers to a small L^2-neighbourhood of the initial image, the absence of a fidelity term inside the inpainting domain K explains the appearance of "spurious edges" in some computations, see Section 3.6. This is also noted in [57], where two numerical realisations of this model for greyscale image inpainting are proposed, one based on the level-set method, and one based on an L^2 flow of an approximation in the Γ-convergence context, like the one we are going to use.

Since treatment of the highest order term (Willmore term) in this energy is particularly challenging, prior works on the L^2 Willmore flow are an important source of inspiration for this work. In particular, we would like to mention the papers [55, 48], where a stable and consistent spatially discrete finite element scheme for the L^2 Willmore flow of a surface, based on its parametric representation, is presented. In [51], a level set formulation for Willmore flow together with spatial and temporal discretisations are proposed. The paper [54] (and others by the same authors) is even closer related to our work. In it, a phase-field approximation is used, and a spatial discretisation of the equilibrium problem and in particular a time-discretisation of the L^2 flow with additional surface and volume constraints is presented, motivated by vesicle membrane deformation models. Our discretisation of the Willmore term employs the same techniques, but is fully discrete. Sources for Willmore flow of curves and its numerical treatment are e.g. [7, 56].

In [4, 5], Ambrosio and Tortorelli introduced a Γ-convergent, elliptic phase-field approximation of the Mumford-Shah functional. For the Mumford-Shah-Euler functional, an analogous approximation for sphere-valued functions is

$$\begin{aligned} E_\varepsilon(\mathbf{u}, s) :=& \frac{\gamma}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2} \int_B |\mathbf{u} - \mathbf{u}_0|^2 \,\mathrm{d}\mathbf{x} \\ &+ \alpha \int_\Omega \left(\frac{\varepsilon}{2} |\nabla s|^2 + \frac{1}{\varepsilon} \psi(s) \right) \mathrm{d}\mathbf{x} + \frac{\beta}{\varepsilon} \int_\Omega \left(\frac{1}{\varepsilon} \psi'(s) - \varepsilon \Delta s \right)^2 \mathrm{d}\mathbf{x}, \end{aligned} \tag{3.1.1}$$

for $\mathbf{u} \in H^1(\Omega, \mathbb{S}^{m-1})$, $\mathbf{u}_0 \in L^\infty(\Omega, \mathbb{S}^{m-1})$, $0 < \varepsilon, k_\varepsilon \ll 1$, $k_\varepsilon = o(\varepsilon)$, and $s \in H^2(\Omega)$. Here, s is a *phase function* approximating $1 - \chi_\Gamma$ by penalisation of phase transitions.

A Γ-convergence result for the curvature and length terms in this functional was conjectured by De Giorgi in [44]. In [16], a proof for the lim sup inequality was given, and in [92] the lim inf inequality, which in this case turns out to be the hard part, could be proved (for the L^1 topology). We are not aware of any corresponding results for the full functional with a sphere-constraint. In [81], the L^2 gradient flow corresponding to the length and curvature terms was studied, and convergence for $\varepsilon \to 0$ was proved using formal asymptotic expansions. Numerical studies for this model (for real-valued functions) were done in [81, 82, 32, 15].

After some preliminaries in Section 3.2, we give the following two numerical algorithms for the L^2 gradient flow of (3.1.1):

3.1.3 Sphere Constraint using Penalisation

In Sections 3.3 and 3.4, a fully discrete, first-order finite element algorithm based on a penalisation of the sphere-constraint using a Ginzburg-Landau term is proposed and analysed; i.e., the L^2 flow of the following energy is considered:

$$\begin{aligned} E_\varepsilon(\mathbf{u}, s) :=& \frac{\gamma}{2} \int_\Omega \left(s^2 + k_\varepsilon\right) |\nabla \mathbf{u}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2} \int_B |\mathbf{u} - \mathbf{u}_0|^2 \,\mathrm{d}\mathbf{x} + \frac{1}{4\delta_\varepsilon} \int_\Omega \left(|\mathbf{u}|^2 - 1\right)^2 \mathrm{d}\mathbf{x} \\ &+ \alpha \int_\Omega \left(\frac{\varepsilon}{2} |\nabla s|^2 + \frac{1}{\varepsilon} \psi(s) \right) \mathrm{d}\mathbf{x} + \frac{\beta}{\varepsilon} \int_\Omega \left(\frac{1}{\varepsilon} \psi'(s) - \varepsilon \Delta s \right)^2 \mathrm{d}\mathbf{x}, \end{aligned}$$

for $\mathbf{u} \in H^1(\Omega, \mathbb{R}^m)$, $\mathbf{u}_0 \in L^\infty(\Omega, \mathbb{R}^m)$, $0 < \varepsilon, \delta_\varepsilon, k_\varepsilon \ll 1$, $k_\varepsilon = o(\varepsilon)$, and $s \in H^2(\Omega)$.

This leads to the following system of equations (formally):

$$\begin{aligned}
\mathbf{u}_t - \gamma \mathrm{div}\left(\left(s^2 + k_\varepsilon\right)\nabla \mathbf{u}\right) + \lambda\chi_B\left(\mathbf{u} - \mathbf{u}_0\right) + \frac{1}{\delta_\varepsilon}\left(|\mathbf{u}|^2 - 1\right)\mathbf{u} &= 0,\\
s_t + \gamma|\nabla \mathbf{u}|^2 s + \frac{\alpha}{\varepsilon}\psi'(s) - \alpha\varepsilon w &\\
+2\beta\left(\frac{1}{\varepsilon^3}\psi''(s)\psi'(s) - \frac{1}{\varepsilon}\Delta\psi'(s) - \frac{1}{\varepsilon^2}\psi''(s)w + \Delta w\right) &= 0,\\
\varepsilon\Delta s - w &= 0,
\end{aligned}$$

with appropriate initial and boundary conditions.

The motivation for this is our experience in [67] and Chapter 2, that penalisation of the sphere-constraint makes it possible to get strong convergence of iterates $\nabla \mathbf{u}_n$ of a splitting scheme of the above system in L^2, while enforcing the sphere-constraint exactly seems to make this impossible. And again this strong convergence proves crucial in identifying limits; i.e., showing that this algorithm actually converges to a weak solution of the L^2 flow.

3.1.4 Sphere Constraint using a Lagrange Multiplier

In Section 3.5, a fully discrete, first-order finite element algorithm using a discrete Lagrange multiplier is proposed and analysed; i.e., the system of equations to be solved looks like this:

$$\begin{aligned}
\mathbf{u}_t - \gamma \mathrm{div}\left(\left(s^2 + k_\varepsilon\right)\nabla \mathbf{u}\right) + \lambda\chi_B\left(\mathbf{u} - \mathbf{u}_0\right) &= \mu\mathbf{u},\\
s_t + \gamma|\nabla \mathbf{u}|^2 s + \frac{\alpha}{\varepsilon}\psi'(s) - \alpha\varepsilon w &\\
+2\beta\left(\frac{1}{\varepsilon^3}\psi''(s)\psi'(s) - \frac{1}{\varepsilon}\Delta\psi'(s) - \frac{1}{\varepsilon^2}\psi''(s)w + \Delta w\right) &= 0,\\
-\varepsilon\Delta s + w &= 0,\\
|\mathbf{u}| &= 1,
\end{aligned}$$

where $\mu : \Omega_T \to \mathbb{R}$ is the Lagrange multiplier for the sphere constraint:

$$\mu = \gamma\left(s^2 + k_\varepsilon\right)|\nabla \mathbf{u}|^2 + \lambda\chi_B\left(1 - \mathbf{u}_0 \cdot \mathbf{u}\right).$$

This approach exactly enforces the sphere constraint. We can show existence of solutions in each step of a discrete splitting scheme, and an energy principle, but we do not get strong convergence of iterates $\nabla \mathbf{u}_n$ in L^2 and therefore cannot identify limits.

Finally, numerical studies are provided in Section 3.6, comparing these two algorithms with each other, with the Mumford-Shah model for inpainting, and with a straightforward channelwise RGB implementation of the Mumford-Shah-Euler model for inpainting.

3.2 Preliminaries

3.2.1 General Notation

We shall use c and C as a generic non-negative constants. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $\langle \mathbf{x}, \mathbf{y} \rangle$ or $\mathbf{x} \cdot \mathbf{y}$ will denote their standard scalar product, and $|\mathbf{x}|$ the Euclidian norm of $\mathbf{x}$. For a set S, $|S|$ denotes its Lebesgue measure of dimension d. The L^2 scalar product and norm will be denoted by $(\cdot, \cdot)$ and $\|\cdot\|$, respectively, and $\mathbb{S}^{m-1}$ will be the unit sphere in $\mathbb{R}^m$.

By $A : B$ for $A, B \in \mathbb{R}^{m\times m}$ we shall denote the dyadic product, i.e., $A : B := \sum_{i,j=1}^m a_{ij} b_{ij}$ for $A = (a_{ij})$, $B = (b_{ij})$. $|A|$ will denote the Frobenius norm of A, i.e., $|A|^2 := \sum_{i,j=1}^m |a_{ij}|^2$. For two vectors $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^m$, let $\mathbf{a} \otimes \mathbf{b} := M$ denote the matrix with entries $m_{ij} := \mathbf{a}_i \mathbf{b}_j$.

We use capital letters for finite element functions and boldface for vectors or vector-valued functions.

3.2.2 Finite Element Space

We shall always assume $\Omega \subset \mathbb{R}^d$ to be a polyhedral Lipschitz domain, $\Omega_T := \Omega \times [0, T]$ for $T \in \mathbb{R}_{>0}$, and $\mathcal{T}_h$ to be a quasi-uniform triangulation of Ω with node set $\mathcal{N}_h$ and maximal mesh size $h > 0$ (c.f. [24]). The space of globally continuous, piecewise affine finite element functions on $\mathcal{T}_h$ is denoted by $V_h(\Omega) \subseteq H^1(\Omega)$. The nodal basis functions are $\{\varphi_{\mathbf{z}} : \mathbf{z} \in \mathcal{N}_h\} \subseteq V_h(\Omega)$. Let $V_h(\Omega, \mathbb{R}^m)$ be the finite element space of $\mathbb{R}^m$-valued mappings with basis functions $\{\boldsymbol{\varphi}^i_{\mathbf{z}} : \mathbf{z} \in \mathcal{N}_h, 1 \le i \le m\}$, with $\boldsymbol{\varphi}^1_{\mathbf{z}} := (\varphi_{\mathbf{z}}, 0, \ldots)^T \in V_h(\Omega, \mathbb{R}^m)$, $\boldsymbol{\varphi}^2_{\mathbf{z}} := (0, \varphi_{\mathbf{z}}, 0, \ldots)^T \in V_h(\Omega, \mathbb{R}^m)$, and so forth. Let $\mathcal{I}_h(\cdot) : C^0\big(\overline{\Omega}\big) \to V_h(\Omega)$ be the Lagrange interpolation operator, and $R_h(\cdot) : H^1(\Omega) \to V_h(\Omega)$ the Ritz projection, defined by

$$\left(\nabla \left(R_h(\varphi) - \varphi\right), \nabla V\right) + \left(R_h(\varphi) - \varphi, V\right) = 0 \quad \forall\, V \in V_h(\Omega)\,, \tag{3.2.1}$$

and let $\boldsymbol{\mathcal{I}}_h(\cdot)$ and $\mathbf{R}_h(\cdot)$ be their vector-valued counterparts. Furthermore, set $(\varphi, \psi)_h := \int_\Omega \mathcal{I}_h(\varphi\psi)\, \mathrm{d}\mathbf{x}$ and $\|\varphi\|_h^2 := (\varphi, \varphi)_h$, and define the discrete Laplace operator $\widetilde{\Delta}_h : V_h(\Omega) \to V_h(\Omega)$ by

$$-\left(\widetilde{\Delta}_h W, V\right)_h = (\nabla W, \nabla V) \quad \forall\, V \in V_h(\Omega)\,. \tag{3.2.2}$$

We remark that

$$\begin{aligned} \|V\| \le \|V\|_h \le (d+2)^{1/2} \|V\|, \text{ and} \\ |(V, W)_h - (V, W)| \le Ch \|V\| \|\nabla W\| \quad \forall\, V, W \in V_h(\Omega)\,. \end{aligned} \tag{3.2.3}$$

Also note the following discrete version of Hölder's inequality: For $\varphi, \psi \in C\big(\overline{\Omega}\big)$, $1 \le p, q \le +\infty$, $\frac{1}{p} + \frac{1}{q} = 1$, and $\beta_{\mathbf{z}} := \int_\Omega \varphi_{\mathbf{z}} \mathrm{d}\mathbf{x}$,

$$\begin{aligned} (\varphi, \psi)_h &= \sum_{\mathbf{z} \in \mathcal{N}_h} \beta_{\mathbf{z}} \varphi(\mathbf{z}) \cdot \psi(\mathbf{z}) \le \left(\sum_{\mathbf{z} \in \mathcal{N}_h} \beta_{\mathbf{z}} |\varphi(\mathbf{z})|^p \right)^{1/p} \left(\sum_{\mathbf{z} \in \mathcal{N}_h} \beta_{\mathbf{z}} |\psi(\mathbf{z})|^q \right)^{1/q} \\ &\le \left\| |\varphi|^{p/2} \right\|_h^{2/p} \left\| |\psi|^{q/2} \right\|_h^{2/q}. \end{aligned}$$

3.2.3 Time-Discretisation

Given a uniform time-discretisation with time-step size $k > 0$, and a sequence $\{\varphi_j\}$ in some Banach space X, we set $d_t\varphi_{j+1} := k^{-1}\{\varphi_{j+1} - \varphi_j\}$ and $\varphi_{j+1/2} := (\varphi_j + \varphi_{j+1})/2$ for $j \geq 0$. Note that $(d_t\varphi_{j+1}, \varphi_{j+1}) = \frac{1}{2}d_t |\varphi_{j+1}|^2 + \frac{k}{2}|d_t\varphi_{j+1}|^2$, if X is a Hilbert space. Piecewise constant interpolations of $\{\varphi_j\}$ are defined for $t \in [jk, (j+1)k)$, and $0 \leq j \leq J-1$ by

$$\varphi_+(t) := \varphi_{j+1}, \qquad \varphi_-(t) := \varphi_j, \qquad \overline{\varphi} := \varphi_{j+1/2},$$

and a piecewise affine interpolation on $[jk, (j+1)k)$ is defined by

$$\varphi(t) := \frac{t - t_j}{k}\varphi_{j+1} + \frac{t_{j+1} - t}{k}\varphi_j.$$

Note that $\partial_t\varphi(t) = \frac{1}{k}(\varphi_{j+1} - \varphi_j) = d_t\varphi_{j+1}$ on $[t_j, t_{j+1})$, and that

$$\left\|\varphi^+ - \varphi\right\|_X + \left\|\overline{\varphi} - \varphi\right\|_X \leq 2k \left\|d_t\varphi\right\|_X. \tag{3.2.4}$$

3.3 Continuous Mumford-Shah-Euler with Sphere Penalisation

In this section, we add a Ginzburg-Landau term penalising the sphere constraint to the functional (3.1.1), so the energy now looks like this:

$$\begin{aligned} E_\varepsilon(\mathbf{u}, s) :=& \frac{\gamma}{2}\int_\Omega \left(s^2 + k_\varepsilon\right)|\nabla\mathbf{u}|^2 \,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_B |\mathbf{u} - \mathbf{u}_0|^2 \,\mathrm{d}\mathbf{x} + \frac{1}{4\delta_\varepsilon}\int_\Omega \left(|\mathbf{u}|^2 - 1\right)^2 \mathrm{d}\mathbf{x} \\ &+ \alpha\int_\Omega \left(\frac{\varepsilon}{2}|\nabla s|^2 + \frac{1}{\varepsilon}\psi(s)\right)\mathrm{d}\mathbf{x} + \frac{\beta}{\varepsilon}\int_\Omega \left(\frac{1}{\varepsilon}\psi'(s) - \varepsilon\Delta s\right)^2 \mathrm{d}\mathbf{x}, \end{aligned} \tag{3.3.1}$$

for $\mathbf{u} \in H^1(\Omega, \mathbb{R}^m)$, $\mathbf{u}_0 \in L^\infty(\Omega, \mathbb{R}^m)$, $0 < \varepsilon, \delta_\varepsilon, k_\varepsilon \ll 1$, $k_\varepsilon = o(\varepsilon)$, and $s \in H^2(\Omega)$.

We shall always assume $\varepsilon, k_\varepsilon, \delta_\varepsilon, \alpha, \beta, \gamma > 0$, and $\lambda \geq 0$ to be fixed.

3.3.1 System of Equations

In order to minimise (3.3.1), we want to formally solve the following system of equations:

$$\begin{aligned} \mathbf{u}_t - \gamma\mathrm{div}\left(\left(s^2 + k_\varepsilon\right)\nabla\mathbf{u}\right) + \lambda\chi_B(\mathbf{u} - \mathbf{u}_0) + \frac{1}{\delta_\varepsilon}\left(|\mathbf{u}|^2 - 1\right)\mathbf{u} &= 0 \text{ in } \Omega_T, \\ s_t + \gamma|\nabla\mathbf{u}|^2 s + \frac{\alpha}{\varepsilon}\psi'(s) - \alpha\varepsilon w & \\ +2\beta\left(\frac{1}{\varepsilon^3}\psi''(s)\psi'(s) - \frac{1}{\varepsilon}\Delta\psi'(s) - \frac{1}{\varepsilon^2}\psi''(s)w + \Delta w\right) &= 0 \text{ in } \Omega_T, \\ \varepsilon\Delta s - w &= 0 \text{ in } \Omega_T, \\ \partial_{\mathbf{n}}\mathbf{u} = \partial_{\mathbf{n}}w = \partial_{\mathbf{n}}s &= 0 \text{ on } \partial\Omega, \text{ and} \\ \mathbf{u}(0,\cdot) = \mathbf{u}_0,\ s(0,\cdot) &= s_0. \end{aligned} \tag{3.3.2}$$

The canonical choice would be $w = -1/\varepsilon\psi'(s) + \varepsilon\Delta s$, but the choice above makes the proof of existence in the discrete setting more manageable (see Step 4 in the proof of Theorem 3.4.2).

3.3.2 Weak Solution

Definition 3.3.1. *Given* $(\mathbf{u}_0, s_0) \in H^1(\Omega, \mathbb{R}^m) \times H^1(\Omega)$*, a tuple* $(\mathbf{u}, s, w)$ *is called a* weak solution *of* (3.3.2)*, if for all* $T > 0$ *it holds that*

(1) $(\mathbf{u}, s, w) \in L^2(0,T; H^1(\Omega, \mathbb{R}^m)) \times L^2(0,T; H^1(\Omega)) \times L^2(0,T; L^2(\Omega))$*, with* $\mathbf{u}(0,\cdot) = \mathbf{u}_0(\cdot)$*, and* $s(0,\cdot) = s_0(\cdot)$ *in the sense of traces, and*

(2) *for all* $\boldsymbol{\varphi} \in C^\infty(\Omega_T, \mathbb{R}^m)$*,* $\varphi \in C^\infty([0,T); C_0^\infty(\Omega))$*, and* $\zeta \in C^\infty(\Omega_T)$*,*

$$
\begin{aligned}
0 = &\int_0^T (\partial_t \mathbf{u}, \boldsymbol{\varphi})\,\mathrm{d}t + \gamma \int_0^T \left(\left(s^2 + k_\varepsilon\right)\nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\right)\mathrm{d}t \\
&+ \lambda \int_0^T \left(\chi_B\left(\mathbf{u} - \mathbf{u}_0\right), \boldsymbol{\varphi}\right)\mathrm{d}t + \frac{1}{\delta_\varepsilon}\int_0^T \left(\left(|\mathbf{u}|^2 - 1\right)\mathbf{u}, \boldsymbol{\varphi}\right)\mathrm{d}t, \\
0 = &\int_0^T (\partial_t s, \varphi)\,\mathrm{d}t + \gamma \int_0^T \left(|\nabla \mathbf{u}|^2 s, \varphi\right)\mathrm{d}t \\
&+ \frac{4\alpha}{\varepsilon}\int_0^T \left(\left(s^2 - 1\right)s, \varphi\right)\mathrm{d}t + \alpha\varepsilon \int_0^T (\nabla s, \nabla \varphi)\,\mathrm{d}t \\
&+ \frac{2\beta}{\varepsilon^3}\int_0^T (\psi''(s)\psi'(s), \varphi)\,\mathrm{d}t + \frac{2\beta}{\varepsilon}\int_0^T (\nabla \psi'(s), \nabla \varphi)\,\mathrm{d}t, \\
&- \frac{2\beta}{\varepsilon^2}\int_0^T (\psi''(s) w, \varphi)\,\mathrm{d}t + 2\beta \int_0^T (w, \Delta \varphi)\,\mathrm{d}t, \\
0 = &\,\varepsilon \int_0^T (\nabla s, \nabla \zeta)\,\mathrm{d}t + \int_0^T (w, \zeta)\,\mathrm{d}t.
\end{aligned}
\tag{3.3.3}
$$

3.3.3 Energy Principle

We proceed formally, by testing $(3.3.2)_1$ with $\mathbf{u}_t$, and $(3.3.2)_{2,3}$ with s_t. We add the two first equations in (3.3.2), and integrate over $[0,T]$:

$$
\begin{aligned}
0 \;=\; &\|\mathbf{u}_t\|^2_{L^2(0,T;L^2)} + \|s_t\|^2_{L^2(0,T;L^2)} + \frac{\gamma}{2}\int_0^T\int_\Omega \partial_t\left(\left(s^2 + k_\varepsilon\right)|\nabla \mathbf{u}|^2\right)\mathrm{d}\mathbf{x}\mathrm{d}t \\
&+\frac{\lambda}{2}\int_0^T\int_B \partial_t\left|\mathbf{u} - \mathbf{u}_0\right|^2\mathrm{d}\mathbf{x}\mathrm{d}t + \frac{1}{4\delta_\varepsilon}\int_0^T\int_\Omega \partial_t\left(|\mathbf{u}|^2 - 1\right)^2\mathrm{d}\mathbf{x}\mathrm{d}t \\
&+\frac{\alpha}{\varepsilon}\int_0^T\int_\Omega \partial_t\psi(s)\mathrm{d}\mathbf{x}\mathrm{d}t + \frac{\alpha\varepsilon}{2}\int_0^T\int_\Omega \partial_t|\nabla s|^2\mathrm{d}\mathbf{x}\mathrm{d}t \\
&+\frac{\beta}{\varepsilon^3}\int_0^T\int_\Omega \partial_t\left|\psi'(s)\right|^2\mathrm{d}\mathbf{x}\mathrm{d}t \\
&+\frac{2\beta}{\varepsilon}\int_0^T \left(\nabla\left(\psi''(s)s_t\right), \nabla s\right) + \left(\nabla\psi'(s), \nabla s_t\right)\mathrm{d}t \\
&+\beta\varepsilon\int_0^T\int_\Omega \partial_t\left|\Delta s\right|^2\mathrm{d}\mathbf{x}\mathrm{d}t.
\end{aligned}
$$

Integration by parts in both terms in the next to last line gives

$$
\frac{2\beta}{\varepsilon}\int_0^T \left(\nabla\left(\psi''(s)s_t\right), \nabla s\right) + \left(\nabla\psi'(s), \nabla s_t\right)\mathrm{d}t = -\frac{2\beta}{\varepsilon}\int_0^T\int_\Omega \partial_t\left(\psi'(s)\Delta s\right)\mathrm{d}\mathbf{x}\mathrm{d}t,
$$

leading to the following energy law:

$$E_\varepsilon(\mathbf{u}(T,\cdot), s(T,\cdot)) + \|\mathbf{u}_t\|^2_{L^2(0,T;L^2)} + \|s_t\|^2_{L^2(0,T;L^2)} = E_\varepsilon(\mathbf{u}(0,\cdot), s(0,\cdot))\,.$$

In the next section, we shall propose a discrete algorithm for the approximate solution of (3.3.2). It will be crucial to ensure an energy principle analogous to the above for this discrete algorithm.

3.4 Discrete Mumford-Shah-Euler with Penalisation

In this section, we shall always assume $\varepsilon, k_\varepsilon, \delta_\varepsilon, \alpha, \beta, \gamma > 0$, and $\lambda \geq 0$ to be fixed. In the context of finite element functions, the energy functional we look at (instead of (3.3.1)), is

$$\begin{aligned} E_{h,\varepsilon}(\mathbf{U}, S) :=& \frac{\gamma}{2}\int_\Omega \left(S^2 + k_\varepsilon\right)|\nabla \mathbf{U}|^2\,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_B |\mathbf{U}-\mathbf{U}_0|^2\,\mathrm{d}\mathbf{x} \\ &+ \frac{1}{4\delta_\varepsilon}\int_\Omega \left(|\mathbf{U}|^2-1\right)^2\mathrm{d}\mathbf{x} + \alpha\int_\Omega \left(\frac{\varepsilon}{2}|\nabla S|^2 + \frac{1}{\varepsilon}\psi(S)\right)\mathrm{d}\mathbf{x} \\ &+ \frac{\beta}{\varepsilon}\int_\Omega \mathcal{I}_h\left(\left(\frac{1}{\varepsilon}\psi'(S) - \varepsilon\widetilde{\Delta}_h S\right)^2\right)\mathrm{d}\mathbf{x}, \end{aligned} \tag{3.4.1}$$

for $\mathbf{U}, \mathbf{U}_0 \in V_h(\Omega, \mathbb{R}^m)$, and $S \in V_h(\Omega)$. Reduced integration in the last term will help us achieve a discrete analog to the energy principle.

3.4.1 Algorithm

In order to get a discrete analog of the energy principle, we use a strategy similar to [54], where a non-stationary time-discretisation of a phase-field approximation to the Willmore flow with surface area and volume constraint is considered.

Algorithm 3.4.1. *Let $\mathbf{U}_0 \in V_h(\Omega, \mathbb{R}^m)$ and $S_0 \in V_h(\Omega)$ be given.*

(1) *Compute $W_0 \in V_h(\Omega)$, such that $(W_0, Z)_h = -\varepsilon\,(\nabla S_0, \nabla Z)$ for all $Z \in V_h(\Omega)$.*

(2) *For $n = 0, \ldots$, compute $S_{n+1}, W_{n+1} \in V_h(\Omega)$, and $\mathbf{U}_{n+1} \in V_h(\Omega, \mathbb{R}^m)$, such that for all $Y, Z \in V_h(\Omega)$ and $\mathbf{X} \in V_h(\Omega, \mathbb{R}^m)$ the following equations hold:*

$$\begin{aligned}
(d_t \mathbf{U}_{n+1}, \mathbf{X}) + \frac{\gamma}{2} \left(\left(S_{n+1}^2 + S_n^2 + 2k_\varepsilon \right) \nabla \mathbf{U}_{n+1/2}, \nabla \mathbf{X} \right) \\
+ \lambda \left(\chi_B \left(\mathbf{U}_{n+1/2} - \mathbf{U}_0 \right), \mathbf{X} \right) \\
+ \frac{1}{2\delta_\varepsilon} \left(\left(|\mathbf{U}_{n+1}|^2 + |\mathbf{U}_n|^2 - 2 \right) \mathbf{U}_{n+1/2}, \mathbf{X} \right) = 0, \\
(d_t S_{n+1}, Y) + \frac{\gamma}{2} \left(\left(|\nabla \mathbf{U}_{n+1}|^2 + |\nabla \mathbf{U}_n|^2 \right) S_{n+1/2}, Y \right) \\
+ \frac{2\alpha}{\varepsilon} \left(\left(S_{n+1}^2 + S_n^2 - 2 \right) S_{n+1/2}, Y \right) + \alpha\varepsilon \left(\nabla S_{n+1/2}, \nabla Y \right) \\
+ \frac{16\beta}{\varepsilon^3} \Big(\left(S_{n+1}^2 + S_{n+1} S_n + S_n^2 - 1 \right) \Big(\left(S_{n+1}^2 - 1 \right) S_{n+1} \\
+ \left(S_n^2 - 1 \right) S_n \Big), Y \Big)_h \\
+ \frac{4\beta}{\varepsilon} \left(\nabla \mathcal{I}_h \big(\left(S_{n+1}^2 - 1 \right) S_{n+1} + \left(S_n^2 - 1 \right) S_n \big), \nabla Y \right) \\
- \frac{8\beta}{\varepsilon^2} \left(\left(S_{n+1}^2 + S_{n+1} S_n + S_n^2 - 1 \right) W_{n+1/2}, Y \right)_h - 2\beta \left(\nabla W_{n+1/2}, \nabla Y \right) = 0, \\
\varepsilon \left(\nabla S_{n+1}, \nabla Z \right) + \left(W_{n+1}, Z \right)_h = 0.
\end{aligned} \tag{3.4.2}$$

We use reduced integration and Lagrange projection in some terms above in order to be able to use non finite element "test functions" in the arguments leading to the energy principle (Step 2 in the proof of Theorem 3.4.2). To solve the above system of equations, we use a simple fixed-point iteration to deal with the coupling of variables and nonlinearities.

3.4.2 Analysis

Theorem 3.4.2. *Let $E_{h,\varepsilon}(\mathbf{U}_0, S_0) \leq C$, independently of h, as well as $\beta > 0$, $d \leq 3$, and $k \leq \widetilde{C}\varepsilon^{-1}h^4$ for sufficiently small $\widetilde{C} \equiv \widetilde{C}(\Omega, \mathcal{T}_h) > 0$ independent of $k, h > 0$. Then Algorithm 3.4.1 converges (up to subsequences) to a weak solution as in Definition 3.3.1.*

We assume $\beta > 0$ for simplicity. The only problematic part with $\beta = 0$ is getting an L^∞ bound on iterates S in space and time, which could e.g. be tackled through additional mass lumping, as described in Section 2.4.

The coupling between time and space discretisation parameters ($k \leq \widetilde{C}\varepsilon^{-1}h^4$) is only needed in the proof of existence of discrete solutions. It arises due to inverse estimates used to control the term $\left(\nabla \mathcal{I}_h \big(\left(S_{n+1}^2 - 1 \right) S_{n+1} + \left(S_n^2 - 1 \right) S_n \big), \nabla Y \right)$, and we assume it can be further improved. For the convergence part of the proof, $k = o(h)$ is sufficient.

In practice, the condition $k \leq \widetilde{C}\varepsilon^{-1}h^4$ is not quite as bad as it looks: While we theoretically assume $\varepsilon > 0$ to be fixed, ε corresponds to the "width" of the interfaces, and we can in practice choose $\varepsilon = ch$, with larger ε making the interfaces more diffuse, but the dynamics more forgiving. In our calculations in Section 3.6, we choose $\varepsilon = 10h$, so $\varepsilon^{-1}h^4$ is practically as good as h^3.

In our simulations, we actually observe that results are reliable (but slow) for $k = h^3$ and unpredictable for $k = h^2$. We therefore use a dynamic time step, described in more detail in Section 3.6.1, that in our examples results in $k \approx 2h^2$ on average, but with much smaller time-steps in the initial phases of the flow.

For identifying limits in the following proof it will be crucial to prove strong L^2 convergence of $\nabla\overline{\mathbf{U}}$ to $\nabla\mathbf{u}$ in space and time, for which we use a strategy derived from [25, Proof of Theorem 2], where the authors show convergence of two adaptive, stationary finite element approximations for the minimisation of the unconstrained Ambrosio-Tortorelli energy: In Step 6 we show that $\mathbf{u}$ fulfils equation $(3.3.3)_1$, then we use equations $(3.3.3)_1$ and $(3.4.2)_1$ and dominated convergence (c.f. Step 1) to show strong $L^2(0,T;L^2(\Omega,\mathbb{R}^m))$ convergence of $\nabla\overline{\mathbf{U}}$ to $\nabla\mathbf{u}$ in Step 7, and finally we use this to show that s fulfils equation $(3.3.3)_2$ in Step 8.

Proof. **Step 1:** *Preliminary Convergence Result.*

Let $p_n, p \in L^\infty(0,T;L^\infty(\Omega))$, such that $\|p_n\|_{L^\infty(0,T;L^\infty)}, \|p\|_{L^\infty(0,T;L^\infty)} \leq c < +\infty$ a.e., independently of n, and $p_n \to p$ in $L^2(0,T;L^2(\Omega))$. Then

$$\lim_n \left\| |p_n - p|^{1/2} |\nabla\boldsymbol{\varphi}| \right\|_{L^2(0,T;L^2)} = 0 \qquad \forall\, \boldsymbol{\varphi} \in L^\infty\big(0,T;H^1(\Omega,\mathbb{R}^m)\big)\,.$$

Proof. Let $\boldsymbol{\varphi} \in L^\infty(0,T;H^1(\Omega,\mathbb{R}^m))$ be fixed. Choose a subsequence $\{p_{n_j}\}_j$ of $\{p_n\}_n$ such that

$$\begin{aligned} \lim_j \left\| \left|p_{n_j} - p\right|^{1/2} |\nabla\boldsymbol{\varphi}| \right\|_{L^2(0,T;L^2)} &= \limsup_n \left\| |p_n - p|^{1/2} |\nabla\boldsymbol{\varphi}| \right\|_{L^2(0,T;L^2)}, \text{ and} \\ \lim_j p_{n_j} &= p \text{ a.e. in } \Omega_T. \end{aligned}$$

Then we have $\left|p_{n_j} - p\right| |\nabla\boldsymbol{\varphi}|^2 \leq 2c\,|\nabla\boldsymbol{\varphi}|^2$ a.e. in Ω_T, by assumption, which leads to $\left\| \left|p_{n_j} - p\right|^{1/2} |\nabla\boldsymbol{\varphi}| \right\|^2_{L^2(\Omega)} \leq c\,\|\nabla\boldsymbol{\varphi}\|^2_{L^2(\Omega)}$ for a.e. $t \in [0,T)$. Therefore, using Lebesgue's Dominated Convergence Theorem twice,

$$\begin{aligned} \limsup_n \left\| |p_n - p|^{1/2} |\nabla\boldsymbol{\varphi}| \right\|^2_{L^2(0,T;L^2)} &= \lim_j \int_0^T \left\| \left|p_{n_j} - p\right|^{1/2} |\nabla\boldsymbol{\varphi}| \right\|^2_{L^2(\Omega)} \mathrm{d}t \\ &= \int_0^T \lim_j \left\| \left|p_{n_j} - p\right|^{1/2} |\nabla\boldsymbol{\varphi}| \right\|^2_{L^2(\Omega)} \mathrm{d}t \\ &= \int_0^T \int_\Omega \lim_j \left|p_{n_j} - p\right| |\nabla\boldsymbol{\varphi}|^2 \,\mathrm{d}\mathbf{x}\mathrm{d}t = 0. \end{aligned}$$

□

Step 2: *Energy principle.*

Let $n \geq 0$, set $\mathbf{X} = d_t\mathbf{U}_{n+1}$ and $Y = d_t S_{n+1}$ in equations $(3.4.2)_1$ and $(3.4.2)_2$, respectively, and multiply both equations with k. For $(3.4.2)_1$, this leads to

$$\begin{aligned} 0 &= k\,\|d_t\mathbf{U}_{n+1}\|^2_{L^2(\Omega)} + \frac{\gamma}{4}\left(S^2_{n+1} + S^2_n + 2k_\varepsilon, |\nabla\mathbf{U}_{n+1}|^2 - |\nabla\mathbf{U}_n|^2\right) \\ &\quad + \frac{\lambda}{2}\left(\chi_B\left(\mathbf{U}_{n+1} + \mathbf{U}_n - 2\mathbf{U}_0\right), \mathbf{U}_{n+1} - \mathbf{U}_n\right) \\ &\quad + \frac{1}{4\delta_\varepsilon}\left(|\mathbf{U}_{n+1}|^2 + |\mathbf{U}_n|^2 - 2, |\mathbf{U}_{n+1}|^2 - |\mathbf{U}_n|^2\right) =: T_1 + T_2 + T_3 + T_4 \end{aligned} \tag{3.4.3}$$

while $(3.4.2)_2$ becomes

$$\begin{aligned}
0 &= k\left\|d_t S_{n+1}\right\|^2_{L^2(\Omega)} + \frac{\gamma}{4}\left(|\nabla \mathbf{U}_{n+1}|^2 + |\nabla \mathbf{U}_n|^2, S_{n+1}^2 - S_n^2\right) \\
&+ \frac{\alpha}{\varepsilon}\left(S_{n+1}^2 + S_n^2 - 2, S_{n+1}^2 - S_n^2\right) + \frac{\alpha\varepsilon}{2}\left(\nabla\left(S_{n+1} + S_n\right), \nabla\left(S_{n+1} - S_n\right)\right) \\
&+ \frac{16\beta}{\varepsilon^3}\Big(\left(S_{n+1}^2 + S_{n+1}S_n + S_n^2 - 1\right)\Big(\left(S_{n+1}^2 - 1\right)S_{n+1} \\
&+ \left(S_n^2 - 1\right)S_n\Big), S_{n+1} - S_n\Big)_h \\
&+ \frac{4\beta}{\varepsilon}\left(\nabla \mathcal{I}_h\big(\left(S_{n+1}^2 - 1\right)S_{n+1} + \left(S_n^2 - 1\right)S_n\big), \nabla\left(S_{n+1} - S_n\right)\right) \\
&- \frac{4\beta}{\varepsilon^2}\left(\left(S_{n+1}^2 + S_{n+1}S_n + S_n^2 - 1\right)\left(W_{n+1} + W_n\right), S_{n+1} - S_n\right)_h \\
&- \beta\left(\nabla\left(W_{n+1} + W_n\right), \nabla\left(S_{n+1} - S_n\right)\right) =: T_5 + \ldots + T_{12}.
\end{aligned} \tag{3.4.4}$$

Thanks to the interpolations introduced in the last four terms, we can now substitute $(3.4.2)_3$, so the last four terms can be rewritten as

$$\begin{aligned}
&\beta^{-1}\left(T_9 + T_{10} + T_{11} + T_{12}\right) \\
&= \frac{16}{\varepsilon^3}\left(\left(S_{n+1}^2 + S_{n+1}S_n + S_n^2 - 1\right)\left(\left(S_{n+1}^2 - 1\right)S_{n+1} + \left(S_n^2 - 1\right)S_n\right), S_{n+1} - S_n\right)_h \\
&\quad + \frac{4}{\varepsilon}\left(\mathcal{I}_h\big(\left(1 - S_{n+1}^2\right)S_{n+1} + \left(1 - S_n^2\right)S_n\big), \widetilde{\Delta}_h\left(S_{n+1} - S_n\right)\right)_h \\
&\quad - \frac{4}{\varepsilon}\left(\left(S_{n+1}^2 + S_{n+1}S_n + S_n^2 - 1\right)\left(S_{n+1} - S_n\right), \widetilde{\Delta}_h\left(S_{n+1} + S_n\right)\right)_h \\
&\quad + \varepsilon\left(\left\|\widetilde{\Delta}_h S_{n+1}\right\|_h^2 - \left\|\widetilde{\Delta}_h S_n\right\|_h^2\right) \\
&= \frac{16}{\varepsilon^3}\left(\left(\left(1 - S_{n+1}^2\right)^2, S_{n+1}^2\right)_h - \left(\left(1 - S_n^2\right)^2, S_n^2\right)_h\right) \\
&\quad + \frac{4}{\varepsilon}\left(\left(1 - S_{n+1}^2\right)S_{n+1} + \left(1 - S_n^2\right)S_n, \widetilde{\Delta}_h\left(S_{n+1} - S_n\right)\right)_h \\
&\quad + \frac{4}{\varepsilon}\left(\left(1 - S_{n+1}^2\right)S_{n+1} - \left(1 - S_n^2\right)S_n, \widetilde{\Delta}_h\left(S_{n+1} + S_n\right)\right)_h \\
&\quad + \varepsilon\left(\left\|\widetilde{\Delta}_h S_{n+1}\right\|_h^2 - \left\|\widetilde{\Delta}_h S_n\right\|_h^2\right) \\
&= \frac{1}{\varepsilon}\int_\Omega \mathcal{I}_h\left(\left(\frac{4}{\varepsilon}\left(1 - S_{n+1}^2\right)S_{n+1} + \varepsilon\widetilde{\Delta}_h S_{n+1}\right)^2 - \left(\frac{4}{\varepsilon}\left(1 - S_n^2\right)S_n + \varepsilon\widetilde{\Delta}_h S_n\right)^2\right)\mathrm{d}\mathbf{x},
\end{aligned}$$

where we used

$$\begin{aligned}
\left(a^2 + ab + b^2 - 1\right)(a - b) &= \left(\left(a^2 - 1\right) + \left(b^2 - 1\right)\right)(a - b) + b\left(a^2 - 1\right) - a\left(b^2 - 1\right) \\
&= a\left(a^2 - 1\right) - b\left(b^2 - 1\right).
\end{aligned}$$

Furthermore

$$T_7 + T_8 = \frac{\alpha}{\varepsilon}\int_\Omega\left(\left(S_{n+1}^2 - 1\right)^2 - \left(S_n^2 - 1\right)^2\right)\mathrm{d}\mathbf{x} + \frac{\alpha\varepsilon}{2}\int_\Omega\left(|\nabla S_{n+1}|^2 - |\nabla S_n|^2\right)\mathrm{d}\mathbf{x},$$

where we used

$$\begin{aligned}(a+b-2)(a-b) &= ((a-1)+(b-1))((a-1)-(b-1)) \\ &= (a-1)^2-(b-1)^2.\end{aligned}$$

Similarly

$$\begin{aligned}T_2+T_6 &= \frac{\gamma}{4}\int_\Omega \left(S_{n+1}^2+S_n^2+2k_\varepsilon\right)\left(|\nabla \mathbf{U}_{n+1}|^2-|\nabla \mathbf{U}_n|^2\right)\mathrm{d}\mathbf{x} \\ &\quad +\frac{\gamma}{4}\int_\Omega \left(S_{n+1}^2-S_n^2\right)\left(|\nabla \mathbf{U}_{n+1}|^2+|\nabla \mathbf{U}_n|^2\right)\mathrm{d}\mathbf{x} \\ &= \frac{\gamma}{2}\int_\Omega \left(\left(S_{n+1}^2+k_\varepsilon\right)|\nabla \mathbf{U}_{n+1}|^2-\left(S_n^2+k_\varepsilon\right)|\nabla \mathbf{U}_n|^2\right)\mathrm{d}\mathbf{x},\end{aligned}$$

and

$$T_4=\frac{1}{4\delta_\varepsilon}\int_\Omega \left(\left(|\mathbf{U}_{n+1}|^2-1\right)^2-\left(|\mathbf{U}_n|^2-1\right)^2\right)\mathrm{d}\mathbf{x},$$

and

$$T_3=\frac{\lambda}{2}\int_\Omega \left(|\mathbf{U}_{n+1}-\mathbf{U}_0|^2-|\mathbf{U}_n-\mathbf{U}_0|^2\right)\mathrm{d}\mathbf{x}.$$

Adding equations (3.4.3) and (3.4.4) therefore leads to

$$E_{h,\varepsilon}(\mathbf{U}_{n+1},S_{n+1})-E_{h,\varepsilon}(\mathbf{U}_n,S_n)=-\frac{1}{k}\|S_{n+1}-S_n\|_{L^2(\Omega)}^2-\frac{1}{k}\|\mathbf{U}_{n+1}-\mathbf{U}_n\|_{L^2(\Omega)}^2, \quad (3.4.5)$$

for any $n\geq 0$.

Step 3: *Uniform boundedness of iterates.*
Thanks to (3.4.5),

$$E_{h,\varepsilon}(\mathbf{U}_N,S_N)+k\sum_{n=0}^{N-1}\|d_t\mathbf{U}_{n+1}\|^2+k\sum_{n=0}^{N-1}\|d_tS_{n+1}\|^2=E_{h,\varepsilon}(\mathbf{U}_0,S_0). \quad (3.4.6)$$

Therefore, using the notation of Section 3.2.3, we immediately get uniform bounds on $\mathbf{U},\overline{\mathbf{U}},\mathbf{U}_+,\mathbf{U}_-$ in $L^\infty(0,T;H^1(\Omega,\mathbb{R}^m))$, on $\partial_t\mathbf{U}$ in $L^2(0,T;L^2(\Omega,\mathbb{R}^m))$, on $S,\overline{S},S_+,S_-$ in $L^\infty(0,T;H^1(\Omega))$, and on $\partial_t S$ in $L^2(0,T;L^2(\Omega))$. In particular, for $d\leq 3$, we have by embedding uniform bounds on $\mathbf{U},\overline{\mathbf{U}},\mathbf{U}_+,\mathbf{U}_-$ in $L^\infty(0,T,L^6(\Omega,\mathbb{R}^m))$, as well as on $S,\overline{S},S_+,S_-$ in $L^\infty(0,T,L^6(\Omega))$.

So, by (3.4.6) and L^p-stability of the Lagrange interpolation,

$$\begin{aligned}\left\|\widetilde{\Delta}_hS_N\right\|_{L^2(\Omega)}^2\leq\left\|\widetilde{\Delta}_hS_N\right\|_h^2 &= \frac{1}{\varepsilon^2}\int_\Omega \mathcal{I}_h\left(\left(\frac{1}{\varepsilon}\psi'(S_N)-\varepsilon\widetilde{\Delta}_hS_N-\frac{1}{\varepsilon}\psi'(S_N)\right)^2\right)\mathrm{d}\mathbf{x} \\ &\leq \frac{c}{\varepsilon^2}\int_\Omega \mathcal{I}_h\left(\left(\frac{1}{\varepsilon}\psi'(S_N)-\varepsilon\widetilde{\Delta}_hS_N\right)^2+\frac{1}{\varepsilon^2}\left(\psi'(S_N)\right)^2\right)\mathrm{d}\mathbf{x} \\ &\leq \frac{c}{\varepsilon}E_{h,\varepsilon}(\mathbf{U}_0,S_0)+\frac{c}{\varepsilon^4}\left\|S_N\left(1-S_N^2\right)\right\|_{L^2(\Omega)}^2.\end{aligned}$$

Let us now define a discrete Laplace operator without mass-lumping, namely $\Delta_h : V_h(\Omega) \to V_h(\Omega)$, with

$$-\left(\Delta_h W, V\right) = \left(\nabla W, \nabla V\right) \quad \forall\, V \in V_h(\Omega)\,.$$

Let $V \in V_h(\Omega)$. Then

$$\begin{aligned}
\left\|\Delta_h V - \widetilde{\Delta}_h V\right\|^2_{L^2(\Omega)} &= \left(\Delta_h V - \widetilde{\Delta}_h V, \Delta_h V - \widetilde{\Delta}_h V\right) \\
&= \left(\Delta_h V, \Delta_h V - \widetilde{\Delta}_h V\right) - \left(\widetilde{\Delta}_h V, \Delta_h V - \widetilde{\Delta}_h V\right)_h \\
&\quad + \left(\widetilde{\Delta}_h V, \Delta_h V - \widetilde{\Delta}_h V\right)_h - \left(\widetilde{\Delta}_h V, \Delta_h V - \widetilde{\Delta}_h V\right) \\
&\leq -\left(\nabla V, \nabla\left(\Delta_h V - \widetilde{\Delta}_h V\right)\right) + \left(\nabla V, \nabla\left(\Delta_h V - \widetilde{\Delta}_h V\right)\right) \\
&\quad + ch \left\|\nabla \widetilde{\Delta}_h V\right\|_{L^2(\Omega)} \left\|\Delta_h V - \widetilde{\Delta}_h V\right\|_{L^2(\Omega)} \\
&\leq c \left\|\widetilde{\Delta}_h V\right\|_{L^2(\Omega)} \left\|\Delta_h V - \widetilde{\Delta}_h V\right\|_{L^2(\Omega)},
\end{aligned}$$

whence

$$\|\Delta_h V\|_{L^2(\Omega)} \leq \left\|\Delta_h V - \widetilde{\Delta}_h V\right\|_{L^2(\Omega)} + \left\|\widetilde{\Delta}_h V\right\|_{L^2(\Omega)} \leq c \left\|\widetilde{\Delta}_h V\right\|_{L^2(\Omega)}.$$

Therefore, using the discrete Sobolev inequalities from [70, Lemma 4.4] and (3.2.3),

$$\begin{aligned}
\|\nabla S_N\|_{L^6(\Omega)} &\leq c\left(\left\|\widetilde{\Delta}_h S_N\right\|_{L^2(\Omega)} + \|\nabla S_N\|_{L^2(\Omega)}\right), \text{ and} \\
\|S_N\|_{L^\infty(\Omega)} + \|\nabla S_N\|_{L^3(\Omega)} &\leq c\,\|\nabla S_N\|^{1/2}_{L^2(\Omega)} \left(\left\|\widetilde{\Delta}_h S_N\right\|^{1/2}_{L^2(\Omega)} + \|\nabla S_N\|^{1/2}_{L^2(\Omega)}\right),
\end{aligned}$$

which ensures uniform bounds on $\nabla S, \nabla \overline{S}, \nabla S_+, \nabla S_-$ in $L^\infty(0,T;L^6(\Omega))$, as well as $S, \overline{S}, S_+, S_-$ in $L^\infty(0,T;L^\infty(\Omega))$.

We also have

$$\begin{aligned}
\|W_N\|^2_{L^2(\Omega)} \leq \|W_N\|^2_h &= \left\|\varepsilon \widetilde{\Delta}_h S_N\right\|^2_h \\
&\leq c\varepsilon E_{h,\varepsilon}(\mathbf{U}_0, S_0) + \frac{c}{\varepsilon^2},
\end{aligned}$$

which ensures uniform bounds on $W, \overline{W}, W_+, W_-$ in $L^\infty(0,T;L^2(\Omega))$.

Furthermore, testing $(3.4.2)_2$ with $Y := W_{n+1/2}$, we can get a better control of $\left\|\nabla W_{n+1/2}\right\|_{L^2(\Omega)}$ than just by inverse estimate, which will be helpful for identifying limits

in the last term of $(3.4.2)_2$:

$$
\begin{aligned}
&2\beta \left\|\nabla W_{n+1/2}\right\|^2_{L^2(\Omega)} \\
&\quad = \left(d_t S_{n+1}, W_{n+1/2}\right) + \frac{\gamma}{2}\left(\left(|\nabla \mathbf{U}_{n+1}|^2 + |\nabla \mathbf{U}_n|^2\right) S_{n+1/2}, W_{n+1/2}\right) \\
&\qquad + \frac{2\alpha}{\varepsilon}\left(\left(S^2_{n+1} + S^2_n - 2\right) S_{n+1/2}, W_{n+1/2}\right) + \alpha\varepsilon\left(\nabla S_{n+1/2}, \nabla W_{n+1/2}\right) \\
&\qquad + \frac{16\beta}{\varepsilon^3}\Big(\left(S^2_{n+1} + S_{n+1}S_n + S^2_n - 1\right)\Big(\left(S^2_{n+1} - 1\right) S_{n+1} \\
&\qquad\quad + \left(S^2_n - 1\right) S_n\Big), W_{n+1/2}\Big)_h \\
&\qquad + \frac{4\beta}{\varepsilon}\left(\nabla \mathcal{I}_h\big(\left(S^2_{n+1} - 1\right) S_{n+1} + \left(S^2_n - 1\right) S_n\big), \nabla W_{n+1/2}\right) \\
&\qquad - \frac{8\beta}{\varepsilon^2}\left(\left(S^2_{n+1} + S_{n+1}S_n + S^2_n - 1\right), W^2_{n+1/2}\right)_h \\
&\quad \le c\left\|d_t S_{n+1}\right\|_{L^2(\Omega)}\left\|W_{n+1/2}\right\|_{L^2(\Omega)} \\
&\qquad + c\left(\left\|\nabla \mathbf{U}_{n+1}\right\|^2_{L^{12/5}(\Omega)} + \left\|\nabla \mathbf{U}_n\right\|^2_{L^{12/5}(\Omega)}\right)\left\|S_{n+1/2}\right\|_{L^\infty(\Omega)}\left\|W_{n+1/2}\right\|_{L^6(\Omega)} \\
&\qquad + \frac{c}{\varepsilon}\left\|S^2_{n+1} + S^2_n - 2\right\|_{L^2(\Omega)}\left\|S_{n+1/2}\right\|_{L^\infty(\Omega)}\left\|W_{n+1/2}\right\|_{L^2(\Omega)} \\
&\qquad + c\varepsilon\left\|\nabla S_{n+1/2}\right\|_{L^2(\Omega)}\left\|\nabla W_{n+1/2}\right\|_{L^2(\Omega)} \\
&\qquad + \frac{c}{\varepsilon^3}\left(\left\|S^5_{n+1}\right\|_{L^\infty(\Omega)} + \left\|S^5_n\right\|_{L^\infty(\Omega)} + 1\right)\left\|W_{n+1/2}\right\|_{L^2(\Omega)} \\
&\qquad + \frac{c}{\varepsilon}\left(\left\|S^2_{n+1}\right\|_{L^\infty(\Omega)}\left\|\nabla S_{n+1}\right\|_{L^2(\Omega)} + \left\|S^2_n\right\|_{L^\infty(\Omega)}\left\|\nabla S_n\right\|_{L^2(\Omega)}\right)\left\|\nabla W_{n+1/2}\right\|_{L^2(\Omega)} \\
&\qquad + \frac{c}{\varepsilon^2}\left(\left\|S_{n+1}\right\|^2_{L^\infty(\Omega)} + \left\|S_n\right\|^2_{L^\infty(\Omega)} + 1\right)\left\|W_{n+1/2}\right\|^2_{L^2(\Omega)} \\
&\quad \le c + \frac{c}{h^{4d/12}}\left(\left\|\nabla \mathbf{U}_{n+1}\right\|^4_{L^2(\Omega)} + \left\|\nabla \mathbf{U}_n\right\|^4_{L^2(\Omega)}\right)\left\|S_{n+1/2}\right\|^2_{L^\infty(\Omega)} + \frac{\beta}{2}\left\|\nabla W_{n+1/2}\right\|^2_{L^2(\Omega)} \\
&\qquad + c\varepsilon^2\left\|\nabla S_{n+1/2}\right\|^2_{L^2(\Omega)} + \frac{\beta}{2}\left\|\nabla W_{n+1/2}\right\|^2_{L^2(\Omega)} \\
&\qquad + \frac{c}{\varepsilon^2}\left(\left\|S^2_{n+1}\right\|^2_{L^\infty(\Omega)}\left\|\nabla S_{n+1}\right\|^2_{L^2(\Omega)} + \left\|S^2_n\right\|^2_{L^\infty(\Omega)}\left\|\nabla S_n\right\|^2_{L^2(\Omega)}\right) + \frac{\beta}{2}\left\|\nabla W_{n+1/2}\right\|^2_{L^2(\Omega)},
\end{aligned}
$$

since $\varepsilon > 0$ constant, using Hölder's and Young's inequalities, Sobolev embedding, the bounds already proved in this step, and the properties of the Lagrange interpolation. Therefore, since $d \le 3$,

$$
\left\|\nabla W_{n+1/2}\right\|^2_{L^2(0,T;L^2)} \le ch^{-1}, \tag{3.4.7}
$$

independently of $k, h > 0$.

Step 4: *Existence of a solution.*

We define a continuous mapping $\mathbf{F} : V_h(\Omega,\mathbb{R}^m) \times V_h(\Omega) \times V_h(\Omega) \to V_h(\Omega,\mathbb{R}^m) \times V_h(\Omega) \times V_h(\Omega)$ by

$$
\begin{aligned}
&(\mathbf{F}(\mathbf{\Phi},\Theta,\Xi),(\mathbf{X},Y,Z))\\
&:= \frac{2}{k}(\mathbf{\Phi}-\mathbf{U}_n,\mathbf{X}) + \frac{\gamma}{2}\left(\left((2\Theta-S_n)^2+S_n^2+2k_\varepsilon\right)\nabla\mathbf{\Phi},\nabla\mathbf{X}\right)\\
&\quad +\lambda\left(\chi_B(\mathbf{\Phi}-\mathbf{U}_0),\mathbf{X}\right)\\
&\quad +\frac{1}{2\delta_\varepsilon}\left(\left(|2\mathbf{\Phi}-\mathbf{U}_n|^2+|\mathbf{U}_n|^2-2\right)\mathbf{\Phi},\mathbf{X}\right)\\
&\quad +\frac{2}{k}(\Theta-S_n,Y) + \frac{\gamma}{2}\left(\left(|\nabla(2\mathbf{\Phi}-\mathbf{U}_n)|^2+|\nabla\mathbf{U}_n|^2\right)\Theta,Y\right)\\
&\quad +\frac{2\alpha}{\varepsilon}\left(\left((2\Theta-S_n)^2+S_n^2-2\right)\Theta,Y\right)+\alpha\varepsilon(\nabla\Theta,\nabla Y)\\
&\quad +\frac{16\beta}{\varepsilon^3}\Big(\left((2\Theta-S_n)^2+(2\Theta-S_n)\cdot S_n+S_n^2-1\right)\left((2\Theta-S_n)^2-1\right)(2\Theta-S_n)\\
&\qquad +\left(S_n^2-1\right)S_n,Y\Big)_h\\
&\quad +\frac{4\beta}{\varepsilon}\left(\nabla\mathcal{I}_h\Big(\left((2\Theta-S_n)^2-1\right)(2\Theta-S_n)+\left(S_n^2-1\right)S_n\Big),\nabla Y\right)\\
&\quad -\frac{8\beta}{\varepsilon^2}\left(\left((2\Theta-S_n)^2+(2\Theta-S_n)\cdot S_n+S_n^2-1\right)\Xi,Y\right)_h-2\beta(\nabla\Xi,\nabla Y)\\
&\quad +2\beta(\nabla\Theta,\nabla Z)+\frac{2\beta}{\varepsilon}(\Xi,Z)_h\\
&=: T_1+\ldots+T_{14}.
\end{aligned}
$$

Now let $(\mathbf{X},Y,Z) := (\mathbf{\Phi},\Theta,\Xi)$. In the following, we repeatedly use Young's and Hölder's inequalities, and standard inverse estimates, see e.g. [24, Section 4.5]. Furthermore, we use the fact that by Step 3 $\|\nabla\mathbf{U}_n\|_{L^2}^2 \le c_0 k_\varepsilon^{-1} = c(E_{h,\varepsilon}(\mathbf{U}_0,S_0))\,k_\varepsilon^{-1}$, $\|\nabla S_n\|_{L^2}^2 \le c_0\varepsilon^{-1}$, as well as $\|S_n\|_{L^\infty} \le c_0\varepsilon^{-5/2}$, $\|\nabla S_n\|_{L^6} \le c_0\varepsilon^{-4}$, and $\|W_n\|_{L^2}^2 \le c_0\varepsilon$.

Note that the cumbersome terms are exactly the ones coming from the curvature term in the energy (terms with coefficient β). Without it, existence would be straightforward.

We now calculate

$$
\begin{aligned}
T_1+T_3 &= \frac{2}{k}(\mathbf{\Phi}-\mathbf{U}_n,\mathbf{\Phi})+\lambda\left(\chi_B(\mathbf{\Phi}-\mathbf{U}_0),\mathbf{\Phi}\right)\\
&\ge \frac{2}{k}\left(\|\mathbf{\Phi}\|_{L^2(\Omega)}^2-\|\mathbf{U}_n\|_{L^2(\Omega)}\|\mathbf{\Phi}\|_{L^2}\right)+\lambda\left(\|\mathbf{\Phi}\|_{L^2(B)}^2-\|\mathbf{\Phi}\|_{L^2(B)}\|\mathbf{U}_0\|_{L^2(B)}\right)\\
&\ge \frac{1}{k}\left(\|\mathbf{\Phi}\|_{L^2(\Omega)}^2-\|\mathbf{U}_n\|_{L^2(\Omega)}^2\right)-\frac{\lambda}{4}\|\mathbf{U}_0\|_{L^2(B)}^2\\
&\ge 0,
\end{aligned}
$$

if $\|\mathbf{\Phi}\|_{L^2} \ge \varrho_1\Big(\|\mathbf{U}_n\|_{L^2},\lambda,\|\mathbf{U}_0\|_{L^2(B)}\Big)$, and

$$
\begin{aligned}
T_2 &= \frac{\gamma}{2}\left(\left((2\Theta-S_n)^2+S_n^2+2k_\varepsilon\right)\nabla\mathbf{\Phi},\nabla\mathbf{\Phi}\right)\\
&\ge \gamma k_\varepsilon\|\nabla\mathbf{\Phi}\|_{L^2}^2 \ge 0.
\end{aligned}
$$

And the next term gives

$$\begin{aligned} T_4 &= \frac{1}{2\delta_\varepsilon}\left(\left(|2\mathbf{\Phi}-\mathbf{U}_n|^2+|\mathbf{U}_n|^2-2\right)\mathbf{\Phi},\mathbf{\Phi}\right) \\ &= \frac{1}{2\delta_\varepsilon}\left(\left(4|\mathbf{\Phi}|^2-4\mathbf{\Phi}\cdot\mathbf{U}_n+2\,|\mathbf{U}_n|^2-2\right)\mathbf{\Phi},\mathbf{\Phi}\right) \\ &\geq \frac{1}{\delta_\varepsilon}\left(|\mathbf{\Phi}|^2-1,|\mathbf{\Phi}|^2\right)>0, \end{aligned}$$

if $\|\mathbf{\Phi}\|_{L^2}\geq\varrho_2(|\Omega|)$.

If we take $\|\Theta\|_{L^2}\geq\varrho_3(\|S_n\|_{L^2})$, and in particular $\|\Theta\|_{L^2}^2\geq 2\,\|S_n\|_{L^2}^2$, then the terms coming from the equation for S give

$$T_5=\frac{2}{k}\left(\Theta-S_n,\Theta\right)\geq\frac{1}{k}\left(\|\Theta\|_{L^2(\Omega)}^2-\|S_n\|_{L^2(\Omega)}^2\right)\geq\frac{1}{2k}\,\|\Theta\|_{L^2(\Omega)}^2>0,$$

which we are going to use below, together with the assumption $k\leq\widetilde{C}\varepsilon^{-1}h^4$. Next

$$T_6 = \frac{\gamma}{2}\left(|\nabla\left(2\mathbf{\Phi}-\mathbf{U}_n\right)|^2+|\nabla\mathbf{U}_n|^2,|\Theta|^2\right)\geq 0,$$

$$\begin{aligned} T_7 &= \frac{2\alpha}{\varepsilon}\left(\left((2\Theta-S_n)^2+S_n^2-2\right),|\Theta|^2\right) \\ &\geq \frac{4\alpha}{\varepsilon}\left(|\Theta|^2-1,|\Theta|^2\right)\geq 0, \end{aligned}$$

if $\|\Theta\|_{L^2}\geq\varrho_4(|\Omega|)$, and

$$T_8=\alpha\varepsilon\,\|\nabla\Theta\|_{L^2(\Omega)}^2\geq 0.$$

And now for the cumbersome terms, the calculations for which we restrict to the leading terms for clarity:

$$\begin{aligned} &\frac{\varepsilon}{2\beta}\left(T_9+T_{10}+T_{11}+T_{12}+T_{13}+T_{14}\right) \\ &\quad=\frac{8}{\varepsilon^2}\Big(\left((2\Theta-S_n)^2+(2\Theta-S_n)\cdot S_n+S_n^2-1\right)\left((2\Theta-S_n)^2-1\right)(2\Theta-S_n) \\ &\qquad\quad+\left(S_n^2-1\right)S_n,\Theta\Big)_h \\ &\qquad+2\left(\nabla\mathcal{I}_h\Big(\left((2\Theta-S_n)^2-1\right)(2\Theta-S_n)+\left(S_n^2-1\right)S_n\Big),\nabla\Theta\right) \\ &\qquad-\frac{4}{\varepsilon}\left(\left((2\Theta-S_n)^2+(2\Theta-S_n)\cdot S_n+S_n^2-1\right)\Xi,\Theta\right)_h+\|\Xi\|_h^2 \\ &\quad\geq\frac{3}{\varepsilon^2}\left\|(2\Theta)^3\right\|_h^2-\frac{c}{h^2}\left\|(2\Theta)^3\right\|_h\|\Theta\|_{L^2(\Omega)}-\frac{2}{\varepsilon}\left\|(2\Theta)^3\right\|_h\|\Xi\|_h+\|\Xi\|_h^2 \\ &\quad\geq\frac{3}{\varepsilon^2}\left\|(2\Theta)^3\right\|_h^2-\frac{1}{\varepsilon^2}\left\|(2\Theta)^3\right\|_h^2-\frac{c\varepsilon^2}{h^4}\|\Theta\|_{L^2(\Omega)}^2-\frac{1}{\varepsilon^2}\left\|(2\Theta)^3\right\|_h^2-\|\Xi\|_h^2+\|\Xi\|_h^2 \\ &\quad\geq\frac{1}{\varepsilon^2}\left\|(2\Theta)^3\right\|_h^2-\frac{c\varepsilon^2}{h^4}\|\Theta\|_{L^2(\Omega)}^2. \end{aligned}$$

Therefore, a term $-c\varepsilon^{-1}h^4\,\|\Theta\|_{L^2(\Omega)}^2$ needs to be compensated by T_5, which can be done since $k\leq\widetilde{C}\varepsilon^{-1}h^4$. So, if $\|\mathbf{\Phi}\|_{L^2(\Omega,\mathbb{R}^m)}\geq\max\{\varrho_1,\varrho_2\}$ and $\|\Theta\|_{L^2(\Omega)}\geq\max\{\varrho_3,\varrho_4\}$ large

enough, then $(\mathbf{F}(\mathbf{\Phi},\Theta,\Xi),(\mathbf{\Phi},\Theta,\Xi))\geq 0$. By a corollary to Brouwer's fixed-point theorem (c.f. [65, p. 279]), there exists a tuple $(\mathbf{U}_{n+1},S_{n+1},W_{n+1})\in V_h(\Omega,\mathbb{R}^m)\times V_h(\Omega)\times V_h(\Omega)$, such that $\mathbf{F}(\mathbf{U}_{n+1},S_{n+1},W_{n+1})=0$, which makes it a solution to (3.4.2).

Step 5: *Convergences* $\{\mathbf{U},S,W\}_{k,h}\to(\mathbf{u},s,w)$.

By Steps 3 and 4, there exist subsequences of $\{\mathbf{U}\}_{k,h},\{\overline{\mathbf{U}}\}_{k,h},\{\mathbf{U}_+\}_{k,h},\{\mathbf{U}_-\}_{k,h}$, and $\{S\}_{k,h},\{\overline{S}\}_{k,h},\{S_+\}_{k,h},\{S_-\}_{k,h}$, and $\{\overline{W}\}_{k,h}$, and mappings $\mathbf{u}:\Omega_T\to\mathbb{R}^m$, and $s,w:\Omega_T\to\mathbb{R}$, such that for $(k,h)\to 0$,

$$\begin{aligned}
\mathbf{U}_+,\mathbf{U}_-,\overline{\mathbf{U}},\mathbf{U}&\overset{*}{\rightharpoonup}\mathbf{u}\text{ in }L^\infty\big(0,T;L^2(\Omega,\mathbb{R}^m)\big),\\
\overline{\mathbf{U}}&\overset{*}{\rightharpoonup}\mathbf{u}\text{ in }L^\infty\big(0,T;H^1(\Omega,\mathbb{R}^m)\big),\\
\partial_t\mathbf{U}&\rightharpoonup\partial_t\mathbf{u}\text{ in }L^2\big(0,T;L^2(\Omega,\mathbb{R}^m)\big),\\
S_+,S_-,\overline{S},S&\overset{*}{\rightharpoonup}s\text{ in }L^\infty\big(0,T;L^2(\Omega)\big),\\
\nabla\overline{S}&\overset{*}{\rightharpoonup}\nabla s\text{ in }L^\infty\big(0,T;L^6\big(\Omega,\mathbb{R}^d\big)\big),\\
\partial_t S&\rightharpoonup\partial_t s\text{ in }L^2\big(0,T;L^2(\Omega)\big),\\
\overline{W}&\overset{*}{\rightharpoonup}w\text{ in }L^\infty\big(0,T;L^2(\Omega)\big).
\end{aligned}\tag{3.4.8}$$

We use (3.2.4) and (3.4.6) to conclude that the different sequences converge to the same limits $\mathbf{u}$ and s:

$$\|\mathbf{U}_*-\mathbf{U}\|^2_{L^2(0,T;L^2)}+\left\|\overline{\mathbf{U}}-\mathbf{U}\right\|^2_{L^2(0,T;L^2)}\leq ck^2\|\partial_t\mathbf{U}\|^2_{L^2(0,T;L^2)}\leq ck^2\overset{\substack{k\to0\\h\to0}}{\longrightarrow}0$$

and

$$\|S_*-S\|^2_{L^2(0,T;L^2)}+\left\|\overline{S}-S\right\|^2_{L^2(0,T;L^2)}\leq ck^2\|\partial_t S\|^2_{L^2(0,T;L^2)}\leq ck^2\overset{\substack{k\to0\\h\to0}}{\longrightarrow}0,$$

for $*\in\{+,-\}$. By an inverse estimate, we get the same for gradients of iterates, if we assume $k=o(h)$, which we shall need in Step 8 below.

Using Aubin-Lions' Lemma, for any $1<q<+\infty$ and any $r<6$ ($d\leq 3$),

$$\begin{aligned}
\mathbf{U}&\to\mathbf{u}\text{ in }L^q(0,T;L^r(\Omega,\mathbb{R}^m)),\\
S&\to s\text{ in }L^q\big(0,T;W^{1,r}(\Omega)\big),
\end{aligned}\tag{3.4.9}$$

Step 6: $\mathbf{u}$ *is part of a weak solution as in Definition 3.3.1.*

Let $\boldsymbol{\varphi}\in C^\infty(\Omega_T,\mathbb{R}^m)$ and $\boldsymbol{\varphi}_h:=\boldsymbol{\mathcal{I}}_h(\boldsymbol{\varphi})$, and assume $\mathbf{u}_0=\mathbf{U}_0$ for simplicity. We rewrite equation $(3.4.2)_1$ as

$$\begin{aligned}
0=&(\partial_t\mathbf{U},\boldsymbol{\varphi}_h)+\frac{\gamma}{2}\left(\left(S_+^2+S_-^2+2k_\varepsilon\right)\nabla\overline{\mathbf{U}},\nabla\boldsymbol{\varphi}_h\right)+\lambda\left(\chi_B\left(\overline{\mathbf{U}}-\mathbf{U}_0\right),\boldsymbol{\varphi}_h\right)\\
&+\frac{1}{2\delta_\varepsilon}\left(\left(|\mathbf{U}_+|^2+|\mathbf{U}_-|^2-2\right)\overline{\mathbf{U}},\boldsymbol{\varphi}_h\right)\\
=:&A_1+\frac{\gamma}{2}A_2+\lambda A_3+\frac{1}{2\delta_\varepsilon}A_4
\end{aligned}\tag{3.4.10}$$

for almost all $t \in [0,T)$. We shall show

$$\begin{aligned} 0 &= \int_0^T (\partial_t \mathbf{u}, \boldsymbol{\varphi})\,\mathrm{d}t + \gamma \int_0^T \left(\left(s^2 + k_\varepsilon\right)\nabla \mathbf{u}, \nabla \boldsymbol{\varphi}\right)\mathrm{d}t + \lambda \int_0^T \left(\chi_B\left(\mathbf{u}-\mathbf{u}_0\right), \boldsymbol{\varphi}\right)\mathrm{d}t \\ &\quad + \frac{1}{\delta_\varepsilon}\int_0^T \left(\left(|\mathbf{u}|^2 - 1\right)\mathbf{u}, \boldsymbol{\varphi}\right)\mathrm{d}t \qquad (3.4.11) \\ &=: \int_0^T \left(B_1 + \frac{\gamma}{2}B_2 + \lambda B_3 + \frac{1}{2\delta_\varepsilon}B_4\right)\mathrm{d}t, \end{aligned}$$

for which we proceed term by term:

$$\begin{aligned} &\left|\int_0^T B_1 \mathrm{d}t - \int_0^T A_1 \mathrm{d}t\right| \\ &\quad \le \left|\int_0^T (\partial_t(\mathbf{u}-\mathbf{U}), \boldsymbol{\varphi})\,\mathrm{d}t\right| + \left|\int_0^T (\partial_t \mathbf{U}, \boldsymbol{\varphi} - \boldsymbol{\varphi}_h)\,\mathrm{d}t\right| \\ &\quad \le \left|\int_0^T (\partial_t(\mathbf{u}-\mathbf{U}), \boldsymbol{\varphi})\,\mathrm{d}t\right| + c\,\|\partial_t \mathbf{U}\|_{L^2(0,T;L^2)}\,\|\boldsymbol{\varphi} - \boldsymbol{\varphi}_h\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0, \end{aligned}$$

by (3.4.8) and the properties of the Lagrange interpolation.

Next,

$$\begin{aligned} &\left|\int_0^T B_2 \mathrm{d}t - \int_0^T A_2 \mathrm{d}t\right| \\ &\quad \le \left|\int_0^T \left(\left(2\left(s^2 - S^2\right) + 2S^2 - S_+^2 - S_-^2\right)\nabla \overline{\mathbf{U}}, \nabla \boldsymbol{\varphi}\right)\mathrm{d}t\right| \\ &\qquad + \left|\int_0^T \left(2\left(s^2 + k_\varepsilon\right)\nabla\left(\mathbf{u} - \overline{\mathbf{U}}\right), \nabla \boldsymbol{\varphi}\right)\mathrm{d}t\right| \\ &\qquad + \left|\int_0^T \left(\left(S_+^2 + S_-^2 + 2k_\varepsilon\right)\nabla \overline{\mathbf{U}}, \nabla\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h\right)\right)\mathrm{d}t\right| \\ &\quad \le c\,\|s - S\|_{L^2(0,T;L^2)}\left\|\nabla \overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)}\|\nabla \boldsymbol{\varphi}\|_{L^\infty(0,T;L^\infty)} \\ &\qquad + ck\,\|\partial_t S\|_{L^2(0,T;L^2)}\left\|\nabla \overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)}\|\nabla \boldsymbol{\varphi}\|_{L^\infty(0,T;L^\infty)} \\ &\qquad + 2\left|\int_0^T \left(\nabla\left(\mathbf{u} - \overline{\mathbf{U}}\right), \left(s^2 + k_\varepsilon\right)\nabla \boldsymbol{\varphi}\right)\mathrm{d}t\right| \\ &\qquad + c\left\|\nabla \overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)}\|\nabla\left(\boldsymbol{\varphi} - \boldsymbol{\varphi}_h\right)\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0, \end{aligned}$$

by (3.4.8), the uniform boundedness of s and its iterates in L^∞ in space and time, and the properties of the Lagrange interpolation.

Next,

$$\begin{aligned}
&\left|\int_0^T B_3\mathrm{d}t - \int_0^T A_3\mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(\chi_B\left(\overline{\mathbf{U}}-\mathbf{U}_0\right),\boldsymbol{\varphi}-\boldsymbol{\varphi}_h\right)\mathrm{d}t\right| + \left|\int_0^T \left(\chi_B\left(\mathbf{u}-\overline{\mathbf{U}}\right),\boldsymbol{\varphi}\right)\mathrm{d}t\right| \\
&\quad\le c\left\|\overline{\mathbf{U}}-\mathbf{U}_0\right\|_{L^2(0,T;L^2)}\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_h\right\|_{L^2(0,T;L^2)} + \left|\int_0^T \left(\chi_B\left(\mathbf{u}-\overline{\mathbf{U}}\right),\boldsymbol{\varphi}\right)\mathrm{d}t\right| \xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}$$

by (3.4.8) and the properties of the Lagrange interpolation.

Next,

$$\begin{aligned}
&\left|\int_0^T B_4 - \int_0^T A_4\mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(\left(|\mathbf{U}_+|^2+|\mathbf{U}_-|^2-2\right)\overline{\mathbf{U}},\boldsymbol{\varphi}-\boldsymbol{\varphi}_h\right)\mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(2\left(|\mathbf{u}|^2-1\right)\mathbf{u} - \left(|\mathbf{U}_+|^2+|\mathbf{U}_-|^2-2\right)\overline{\mathbf{U}},\boldsymbol{\varphi}\right)\mathrm{d}t\right| \\
&\quad\le c\left(\|\mathbf{U}_+\|^2_{L^4(0,T;L^4)}+1\right)\left\|\overline{\mathbf{U}}\right\|_{L^4(0,T;L^4)}\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_h\right\|_{L^4(0,T;L^4)} \\
&\qquad + c\left(\|\mathbf{U}_-\|^2_{L^4(0,T;L^4)}+1\right)\left\|\overline{\mathbf{U}}\right\|_{L^4(0,T;L^4)}\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_h\right\|_{L^4(0,T;L^4)} \\
&\qquad + \left|\int_0^T \left(2\left(|\mathbf{u}|^2-1\right)\left(\mathbf{u}-\mathbf{U}+\mathbf{U}-\overline{\mathbf{U}}\right),\boldsymbol{\varphi}\right)\mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(\left(2\left(|\mathbf{u}|^2-|\mathbf{U}|^2\right)+2|\mathbf{U}|^2-|\mathbf{U}_+|^2-|\mathbf{U}_-|^2\right)\overline{\mathbf{U}},\boldsymbol{\varphi}\right)\mathrm{d}t\right| \\
&\quad\le c\left(\|\mathbf{U}_+\|^2_{L^4(0,T;L^4)}+1\right)\left\|\overline{\mathbf{U}}\right\|_{L^4(0,T;L^4)}\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_h\right\|_{L^4(0,T;L^4)} \\
&\qquad + c\left(\|\mathbf{U}_-\|^2_{L^4(0,T;L^4)}+1\right)\left\|\overline{\mathbf{U}}\right\|_{L^4(0,T;L^4)}\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}_h\right\|_{L^4(0,T;L^4)} \\
&\qquad + c\left(\|\mathbf{u}\|^2_{L^4(0,T;L^4)}+1\right)\|\mathbf{u}-\mathbf{U}\|_{L^2(0,T;L^2)}\|\boldsymbol{\varphi}\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + c\left(\|\mathbf{u}\|^2_{L^4(0,T;L^4)}+1\right)k\,\|\partial_t\mathbf{U}\|_{L^2(0,T;L^2)}\|\boldsymbol{\varphi}\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + c\,\|\mathbf{u}-\mathbf{U}\|_{L^2(0,T;L^2)}\|\mathbf{u}+\mathbf{U}\|_{L^2(0,T;L^2)}\left\|\overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)}\|\boldsymbol{\varphi}\|_{L^2(0,T;L^2)} \\
&\qquad + ck\,\|\partial_t\mathbf{U}\|_{L^2(0,T;L^2)}\|\mathbf{U}+\mathbf{U}_+\|_{L^2(0,T;L^2)}\left\|\overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)}\|\boldsymbol{\varphi}\|_{L^2(0,T;L^2)} \\
&\qquad + ck\,\|\partial_t\mathbf{U}\|_{L^2(0,T;L^2)}\|\mathbf{U}+\mathbf{U}_-\|_{L^2(0,T;L^2)}\left\|\overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)}\|\boldsymbol{\varphi}\|_{L^2(0,T;L^2)} \\
&\quad\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}$$

by (3.2.4), (3.4.8), (3.4.9), and the properties of the Lagrange interpolation.

This concludes the proof of (3.4.11).

Step 7 : $\nabla\overline{\mathbf{U}} \to \nabla\mathbf{u}$ *strongly in* $L^2\big(0,T;L^2\big(\Omega,\mathbb{R}^{m\times d}\big)\big)$.

Let $\mathbf{u}_h := \mathbf{R}_h(\mathbf{u})$ and put $\boldsymbol{\varphi} := \boldsymbol{\varphi}_h := \overline{\mathbf{U}} - \mathbf{u}_h$ into (3.4.10) and (3.4.11). Subtracting the former from the latter, we get

$$\begin{aligned}
&\frac{\gamma}{2}\int_0^T \left(\left(S_+^2 + S_-^2 + 2k_\varepsilon\right)\nabla\left(\overline{\mathbf{U}} - \mathbf{u}\right), \nabla\left(\overline{\mathbf{U}} - \mathbf{u}_h\right)\right)\mathrm{d}t \\
&\quad= \frac{\gamma}{2}\int_0^T \left(\left(2s^2 - S_+^2 - S_-^2\right)\nabla\mathbf{u}, \nabla\left(\overline{\mathbf{U}} - \mathbf{u}_h\right)\right)\mathrm{d}t \\
&\qquad+ \int_0^T \left(\partial_t\left(\mathbf{u} - \mathbf{U}\right), \overline{\mathbf{U}} - \mathbf{u}_h\right)\mathrm{d}t + \lambda\int_0^T \left(\chi_B\left(\mathbf{u} - \overline{\mathbf{U}}\right), \overline{\mathbf{U}} - \mathbf{u}_h\right)\mathrm{d}t \\
&\qquad+ \frac{1}{2\delta_\varepsilon}\int_0^T \left(2\left(|\mathbf{u}|^2 - 1\right)\mathbf{u} - \left(|\mathbf{U}_+|^2 + |\mathbf{U}_-|^2 - 2\right)\overline{\mathbf{U}}, \overline{\mathbf{U}} - \mathbf{u}_h\right)\mathrm{d}t.
\end{aligned}$$

We therefore calculate

$$\begin{aligned}
&\gamma k_\varepsilon \left\|\nabla\left(\overline{\mathbf{U}} - \mathbf{u}\right)\right\|_{L^2(0,T;L^2)}^2 \\
&\quad\le \frac{\gamma}{2}\int_0^T \left(\left(S_+^2 + S_-^2 + 2k_\varepsilon\right)\nabla\left(\overline{\mathbf{U}} - \mathbf{u}\right), \nabla\left(\overline{\mathbf{U}} - \mathbf{u}_h\right)\right)\mathrm{d}t \\
&\qquad+ \frac{\gamma}{2}\int_0^T \left(\left(S_+^2 + S_-^2 + 2k_\varepsilon\right)\nabla\left(\overline{\mathbf{U}} - \mathbf{u}\right), \nabla\left(\mathbf{u}_h - \mathbf{u}\right)\right)\mathrm{d}t \\
&\quad= \frac{\gamma}{2}\int_0^T \left(\left(S_+^2 + S_-^2 + 2k_\varepsilon\right)\nabla\left(\overline{\mathbf{U}} - \mathbf{u}\right), \nabla\left(\mathbf{u}_h - \mathbf{u}\right)\right)\mathrm{d}t \\
&\qquad+ \frac{\gamma}{2}\int_0^T \left(\left(2s^2 - S_+^2 - S_-^2\right)\nabla\mathbf{u}, \nabla\left(\overline{\mathbf{U}} - \mathbf{u}_h\right)\right)\mathrm{d}t \\
&\qquad+ \int_0^T \left(\partial_t\left(\mathbf{u} - \mathbf{U}\right), \overline{\mathbf{U}} - \mathbf{u}_h\right)\mathrm{d}t + \lambda\int_0^T \left(\chi_B\left(\mathbf{u} - \overline{\mathbf{U}}\right), \overline{\mathbf{U}} - \mathbf{u}_h\right)\mathrm{d}t \\
&\qquad+ \frac{1}{2\delta_\varepsilon}\int_0^T \left(2\left(|\mathbf{u}|^2 - 1\right)\mathbf{u} - \left(|\mathbf{U}_+|^2 + |\mathbf{U}_-|^2 - 2\right)\overline{\mathbf{U}}, \overline{\mathbf{U}} - \mathbf{u}_h\right)\mathrm{d}t \\
&\quad=: \frac{\gamma}{2}T_1 + \frac{\gamma}{2}T_2 + T_3 + \lambda T_4 + \frac{1}{2\delta_\varepsilon}T_5.
\end{aligned}$$

We now show that in the limit $(k, h) \to 0$, each of these terms vanishes, noting first that $\nabla\left(\mathbf{u}_h - \mathbf{u}\right)$ and $\overline{\mathbf{U}} - \mathbf{u}_h$ converge strongly in $L^2(0,T;L^2)$:

$$\left\|\nabla\left(\mathbf{u}_h - \mathbf{u}\right)\right\|_{L^2(0,T;L^2)} = \left\|\nabla\left(\mathbf{R}_h(\mathbf{u}) - \mathbf{u}\right)\right\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0, \tag{3.4.12}$$

by a density argument, and

$$\left\|\overline{\mathbf{U}} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \le \left\|\overline{\mathbf{U}} - \mathbf{u}\right\|_{L^2(0,T;L^2)} + \left\|\mathbf{u} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0,$$

by (3.4.8), (3.4.9), and (3.4.12). Therefore,

$$\begin{aligned}
T_1 &\le c\left\|S_+^2 + S_-^2 + 2k_\varepsilon\right\|_{L^\infty(0,T;L^\infty)}\left\|\nabla\left(\overline{\mathbf{U}} - \mathbf{u}\right)\right\|_{L^2(0,T;L^2)}\left\|\nabla\left(\mathbf{u}_h - \mathbf{u}\right)\right\|_{L^2(0,T;L^2)} \\
&\le c\left\|\nabla\left(\mathbf{u}_h - \mathbf{u}\right)\right\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}$$

and

$$\begin{aligned} T_2 &\leq \frac{\gamma}{2}\int_0^T \left(\left(2s^2 - S_+^2 - S_-^2\right)\nabla\mathbf{u}, \nabla\left(\overline{\mathbf{U}} - \mathbf{u}_h\right)\right)\mathrm{d}t \\ &\leq c\left\| |s-S_*|^{1/2}\,|\nabla\mathbf{u}|\right\|_{L^2(0,T;L^2)} \left\|\nabla\left(\overline{\mathbf{U}} - \mathbf{u}_h\right)\right\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0, \end{aligned}$$

by the uniform boundedness of s and its iterates in L^∞ in space and time, and Step 1; where we sum over $* \in \{+,-\}$. Furthermore,

$$T_3 \leq c\left\|\partial_t\left(\mathbf{u}-\mathbf{U}\right)\right\|_{L^2(0,T;L^2)} \left\|\overline{\mathbf{U}} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0,$$

and

$$T_4 \leq c\left\|\mathbf{u} - \overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)} \left\|\overline{\mathbf{U}} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0.$$

Finally, the penalty term tends to zero for $(k,h)\to 0$ since

$$\begin{aligned} T_5 &\leq c\left\|2\left(|\mathbf{u}|^2-1\right)\mathbf{u} - \left(|\mathbf{U}_+|^2 + |\mathbf{U}_-|^2 - 2\right)\overline{\mathbf{U}}\right\|_{L^2(0,T;L^2)} \left\|\overline{\mathbf{U}} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \\ &\leq c\left(\|\mathbf{u}\|^3_{L^6(0,T;L^6)} + \|\mathbf{u}\|_{L^2(0,T;L^2)}\right)\left\|\overline{\mathbf{U}} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \\ &\quad + c\left(\|\mathbf{U}_+\|^2_{L^6(0,T;L^6)} + \|\mathbf{U}_-\|^2_{L^6(0,T;L^6)} + 2\right)\left\|\overline{\mathbf{U}}\right\|_{L^6(0,T;L^6)} \left\|\overline{\mathbf{U}} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \\ &\leq c\left\|\overline{\mathbf{U}} - \mathbf{u}_h\right\|_{L^2(0,T;L^2)} \xrightarrow[h\to 0]{k\to 0} 0, \end{aligned}$$

as claimed.

Step 8: *s is part of a weak solution as in Definition 3.3.1.*

Let $\varphi \in C_0^\infty(\Omega_T)$, $\varphi_h := \mathcal{I}_h(\varphi)$, and $\psi \in C^\infty(\Omega_T)$, $\psi_h := \mathcal{I}_h(\psi)$.

We rewrite equations $(3.4.2)_{2,3}$ (adding up two consecutive versions of $(3.4.2)_3$): For almost all $t \in [0,T)$ it holds that

$$\begin{aligned} 0 &= (\partial_t S, \varphi_h) + \frac{\gamma}{2}\left(\left(|\nabla\mathbf{U}_+|^2 + |\nabla\mathbf{U}_-|^2\right)\overline{S}, \varphi_h\right) \\ &\quad + \frac{2\alpha}{\varepsilon}\left(\left(S_+^2 + S_-^2 - 2\right)\overline{S}, \varphi_h\right) + \alpha\varepsilon\left(\nabla\overline{S}, \nabla\varphi_h\right) \\ &\quad + \frac{16\beta}{\varepsilon^3}\left(\left(S_+^2 + S_+S_- + S_-^2 - 1\right)\left(\left(S_+^2-1\right)S_+ + \left(S_-^2-1\right)S_-\right), \varphi_h\right)_h \\ &\quad + \frac{4\beta}{\varepsilon}\left(\nabla\mathcal{I}_h\left(\left(S_+^2-1\right)S_+ + \left(S_-^2-1\right)S_-\right), \nabla\varphi_h\right) \\ &\quad - \frac{8\beta}{\varepsilon^2}\left(\left(S_+^2 + S_+S_- + S_-^2 - 1\right)\overline{W}, \varphi_h\right)_h + 2\beta\left(\overline{W}, \widetilde{\Delta}_h\varphi_h\right)_h \\ &=: A_1 + \frac{\gamma}{2}A_2 + \frac{2\alpha}{\varepsilon}A_3 + \alpha\varepsilon A_4 + \frac{16\beta}{\varepsilon^3}A_5 + \frac{4\beta}{\varepsilon}A_6 + \frac{8\beta}{\varepsilon^2}A_7 + 2\beta A_8, \text{ and} \\ 0 &= \varepsilon\left(\nabla\overline{S}, \nabla\psi_h\right) + \left(\overline{W}, \psi_h\right)_h \\ &=: \varepsilon A_9 + A_{10}. \end{aligned} \tag{3.4.13}$$

We need to prove the following two equations:

$$
\begin{aligned}
0 &= \int_0^T (\partial_t s, \varphi)\,\mathrm{d}t + \gamma \int_0^T \left(|\nabla \mathbf{u}|^2 s, \varphi\right)\mathrm{d}t \\
&\quad + \frac{4\alpha}{\varepsilon}\int_0^T \left(\left(s^2-1\right)s, \varphi\right)\mathrm{d}t + \alpha\varepsilon \int_0^T (\nabla s, \nabla\varphi)\,\mathrm{d}t \\
&\quad + \frac{2\beta}{\varepsilon^3}\int_0^T (\psi''(s)\psi'(s), \varphi)\,\mathrm{d}t + \frac{2\beta}{\varepsilon}\int_0^T (\nabla\psi'(s), \nabla\varphi)\,\mathrm{d}t \\
&\quad - \frac{2\beta}{\varepsilon^2}\int_0^T (\psi''(s)w, \varphi)\,\mathrm{d}t + 2\beta\int_0^T (w, \Delta\varphi)\,\mathrm{d}t \\
&=: \int_0^T \left(B_1 + \frac{\gamma}{2}B_2 + \frac{2\alpha}{\varepsilon}B_3 + \alpha\varepsilon B_4\right)\mathrm{d}t \\
&\quad + \int_0^T \left(\frac{16\beta}{\varepsilon^3}B_5 + \frac{4\beta}{\varepsilon}B_6 + \frac{8\beta}{\varepsilon^2}B_7 + 2\beta B_8\right)\mathrm{d}t, \text{ and} \\
0 &= \varepsilon\int_0^T (\nabla s, \nabla\psi)\,\mathrm{d}t + \int_0^T (w, \psi)\,\mathrm{d}t \\
&=: \int_0^T (\varepsilon B_9 + B_{10})\,\mathrm{d}t,
\end{aligned}
\tag{3.4.14}
$$

for which we proceed term by term:

$$
\begin{aligned}
&\left|\int_0^T B_1 \mathrm{d}t - \int_0^T A_1 \mathrm{d}t\right| \\
&\quad \le \left|\int_0^T (\partial_t S, \varphi - \varphi_h)\,\mathrm{d}t\right| + \left|\int_0^T (\partial_t(s-S), \varphi)\,\mathrm{d}t\right| \\
&\quad \le c\,\|\partial_t S\|_{L^2(0,T;L^2)}\,\|\varphi - \varphi_h\|_{L^2(0,T;L^2)} + \left|\int_0^T (\partial_t(s-S), \varphi)\,\mathrm{d}t\right| \\
&\quad \xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}
$$

by (3.4.8) and the properties of the Lagrange interpolation.

$$
\begin{aligned}
&\left|\int_0^T B_2 - \int_0^T A_2 \mathrm{d}t\right| \mathrm{d}t \\
&\quad\le \left|\int_0^T \left(\left(|\nabla \mathbf{U}_+|^2 + |\nabla \mathbf{U}_-|^2\right)\overline{S}, \varphi - \varphi_h\right) \mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(|\nabla \mathbf{u}|^2 s - |\nabla \mathbf{U}_+|^2\,\overline{S}, \varphi\right) \mathrm{d}t\right| + \left|\int_0^T \left(|\nabla \mathbf{u}|^2 s - |\nabla \mathbf{U}_-|^2\,\overline{S}, \varphi\right) \mathrm{d}t\right| \\
&\quad\le c \left\||\nabla \mathbf{U}_+|^2 + |\nabla \mathbf{U}_-|^2\right\|_{L^1(0,T;L^1)} \left\|\overline{S}\right\|_{L^\infty(0,T;L^\infty)} \|\varphi - \varphi_h\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + \left|\int_0^T \left(\left(|\nabla \mathbf{u}|^2 - |\nabla \mathbf{U}_+|^2\right)\overline{S}, \varphi\right) \mathrm{d}t\right| + \left|\int_0^T \left(\left(|\nabla \mathbf{u}|^2 - |\nabla \mathbf{U}_-|^2\right)\overline{S}, \varphi\right) \mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(2|\nabla \mathbf{u}|^2 \left(s - \overline{S}\right), \varphi\right) \mathrm{d}t\right| \\
&\quad\le c \left\||\nabla \mathbf{U}_+|^2 + |\nabla \mathbf{U}_-|^2\right\|_{L^1(0,T;L^1)} \left\|\overline{S}\right\|_{L^\infty(0,T;L^\infty)} \|\varphi - \varphi_h\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + c \|\nabla \mathbf{U}_* - \nabla \mathbf{u}\|_{L^2(0,T;L^2)} \|\nabla \mathbf{U}_* + \nabla \mathbf{u}\|_{L^2(0,T;L^2)} \left\|\overline{S}\right\|_{L^\infty(0,T;L^\infty)} \|\varphi\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + c \left\|\left(s - \overline{S}\right)|\nabla \mathbf{u}|^2\right\|_{L^1(0,T;L^1)} \|\varphi\|_{L^\infty(0,T;L^\infty)} \\
&\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}
$$

where we sum over $* \in \{+,-\}$ and use

$$
\|\mathbf{U}_* - \mathbf{U}\|_{H^1(\Omega)} + \left\|\overline{\mathbf{U}} - \mathbf{U}\right\|_{H^1(\Omega)} \le c\frac{k}{h} \|\partial_t \mathbf{U}\|_{L^2(\Omega)} \le c\frac{k}{h} \xrightarrow[h\to 0]{k\to 0} 0
$$

— by $k = o(h)$ and an inverse estimate —, as well as (3.4.8), Step 1, and the properties of the Lagrange interpolation.

Next,

$$
\begin{aligned}
&\left|\int_0^T B_3 \mathrm{d}t - \int_0^T A_3 \mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(\left(S_+^2 + S_-^2 - 2\right)\overline{S}, \varphi - \varphi_h\right) \mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(\left(s^2 - 1\right) s - \left(S_+^2 - 1\right)\overline{S}, \varphi\right) \mathrm{d}t\right| + \left|\int_0^T \left(\left(s^2 - 1\right) s - \left(S_-^2 - 1\right)\overline{S}, \varphi\right) \mathrm{d}t\right| \\
&\quad\le \left\|S_+^2 + S_-^2 - 2\right\|_{L^4(0,T;L^4)}^2 \left\|\overline{S}\right\|_{L^2(0,T;L^2)} \|\varphi - \varphi_h\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + c \|s - S_*\|_{L^2(0,T;L^2)} \|s + S_*\|_{L^\infty(0,T;L^\infty)} \|s\|_{L^\infty(0,T;L^\infty)} \|\varphi\|_{L^2(0,T;L^2)} \\
&\qquad + c \left\|S_*^2 - 1\right\|_{L^\infty(0,T;L^\infty)} \left\|s - \overline{S}\right\|_{L^2(0,T;L^2)} \|\varphi\|_{L^2(0,T;L^2)} \\
&\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}
$$

by (3.4.8) and the properties of the Lagrange interpolation; where we sum over $* \in \{+,-\}$.

Next,

$$\begin{aligned}
&\left|\int_0^T B_4 \mathrm{d}t - \int_0^T A_4 \mathrm{d}t\right| \\
&\quad \leq \left|\int_0^T \left(\nabla \overline{S}, \nabla \varphi_h\right) \mathrm{d}t\right| + \left|\int_0^T \left(\nabla \overline{S}, \nabla \left(\varphi - \varphi_h\right)\right) \mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(\nabla \left(s - \overline{S}\right), \nabla \varphi\right) \mathrm{d}t\right| \\
&\quad \leq c \left\|\nabla \overline{S}\right\|_{L^2(0,T;L^2)} \left\|\nabla \left(\varphi - \varphi_h\right)\right\|_{L^2(0,T;L^2)} + \left|\int_0^T \left(\nabla \left(s - \overline{S}\right), \nabla \varphi\right) \mathrm{d}t\right| \\
&\xrightarrow[h \to 0]{k \to 0} 0,
\end{aligned}$$

by (3.4.8) and the properties of the Lagrange interpolation.

For convenience, we now introduce the following notation:

$$\begin{aligned}
F &:= S_+^2 + S_+ S_- + S_-^2 - 1, \\
f &:= 3s^2 - 1, \\
G_1 &:= \left(S_+^2 - 1\right) S_+, \\
G_2 &:= \left(S_-^2 - 1\right) S_-, \\
g &:= \left(s^2 - 1\right) s,
\end{aligned}$$

all of which are uniformly bounded in $L^\infty(0,T;L^\infty(\Omega))$ and $L^\infty(0,T;W^{1,6}(\Omega))$. Now note that, always summing over $* \in \{+,-\}$,

$$\left\|S_* - S\right\|_{H^1(\Omega)} + \left\|\overline{S} - S\right\|_{H^1(\Omega)} \leq c \frac{k}{h} \left\|\partial_t S\right\|_{L^2(\Omega)}^2 \leq c \frac{k}{h} \xrightarrow[h \to 0]{k \to 0} 0, \tag{3.4.15}$$

by $k = o(h)$ and an inverse estimate, whence, with $\left\|S_* - s\right\|_{H^1(\Omega)} \to 0$,

$$\begin{aligned}
\left\|f - F\right\|_{L^2(0,T;H^1)} &\leq c \left\|s - S_*\right\|_{L^4(0,T;L^4)} \left\|s + S_*\right\|_{L^\infty(0,T;L^\infty)} \left\|\nabla S_*\right\|_{L^4(0,T;L^4)} \\
&\quad + c \left\|s - S_*\right\|_{L^2(0,T;H^1)} \left\|s^2\right\|_{L^\infty(0,T;L^\infty)} \xrightarrow[h \to 0]{k \to 0} 0,
\end{aligned} \tag{3.4.16}$$

and analogously,

$$\left\|2g - (G_1 + G_2)\right\|_{L^2(0,T;H^1)} \xrightarrow[h \to 0]{k \to 0} 0. \tag{3.4.17}$$

Therefore the next term gives

$$
\begin{aligned}
&\left|\int_0^T B_5 \mathrm{d}t - \int_0^T A_5 \mathrm{d}t\right| \\
&\quad \le \left|\int_0^T \left(f\left(2g - (G_1+G_2)\right), \varphi\right) \mathrm{d}t\right| + \left|\int_0^T \left(\left(G_1+G_2\right)(f-F), \varphi\right) \mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(\left(G_1+G_2\right)F, \varphi-\varphi_h\right) \mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(\left(G_1+G_2\right)F - R_h\left(\left(G_1+G_2\right)F\right), \varphi_h\right) \mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(R_h\left(\left(G_1+G_2\right)F\right), \varphi_h\right) - \left(R_h\left(\left(G_1+G_2\right)F\right), \varphi_h\right)_h \mathrm{d}t\right| \\
&\quad \le c\left\|f\right\|_{L^\infty(0,T;L^\infty)} \left\|2g - (G_1+G_2)\right\|_{L^1(0,T;L^1)} \left\|\varphi\right\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + c\left\|G_1+G_2\right\|_{L^\infty(0,T;L^\infty)} \left\|f-F\right\|_{L^1(0,T;L^1)} \left\|\varphi\right\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + c\left\|G_1+G_2\right\|_{L^\infty(0,T;L^\infty)} \left\|F\right\|_{L^\infty(0,T;L^\infty)} \left\|\varphi-\varphi_h\right\|_{L^1(0,T;L^1)} \\
&\qquad + ch\left\|\varphi_h\right\|_{L^\infty(0,T;L^\infty)} \left\|\nabla\left(G_1+G_2\right)\right\|_{L^2(0,T;L^2)} \left\|F\right\|_{L^2(0,T;L^2)} \\
&\qquad + ch\left\|\varphi_h\right\|_{L^\infty(0,T;L^\infty)} \left\|\nabla F\right\|_{L^2(0,T;L^2)} \left\|G_1+G_2\right\|_{L^2(0,T;L^2)} \\
&\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}
$$

by (3.4.8), (3.2.3), (3.4.16), (3.4.17), and the properties of the Ritz projection and the Lagrange interpolation.

Next,

$$
\begin{aligned}
&\left|\int_0^T B_6 \mathrm{d}t - \int_0^T A_6 \mathrm{d}t\right| \\
&\quad \le \left|\int_0^T \left(\nabla\left(2g - \mathcal{I}_h(G_1+G_2)\right), \nabla\varphi\right) \mathrm{d}t\right| + \left|\int_0^T \left(\nabla\mathcal{I}_h(G_1+G_2), \nabla\left(\varphi-\varphi_h\right)\right) \mathrm{d}t\right| \\
&\quad \le c\left\|2g - (G_1+G_2)\right\|_{L^2(0,T;L^2)} \left\|\varphi\right\|_{L^2(0,T;H^2)} \\
&\qquad + c\left\|(G_1+G_2) - \mathcal{I}_h(G_1+G_2)\right\|_{L^2(0,T;L^2)} \left\|\varphi\right\|_{L^2(0,T;H^2)} \\
&\qquad + c\left\|G_1+G_2\right\|_{L^2(0,T;H^1)} \left\|\varphi-\varphi_h\right\|_{L^2(0,T;H^1)} \\
&\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}
$$

by (3.4.8), (3.4.16), (3.4.17), and the properties of the Lagrange interpolation.

Next,

$$\begin{aligned}
&\left|\int_0^T B_7\mathrm{d}t - \int_0^T A_7\mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(w-\overline{W}, f\varphi\right)\mathrm{d}t\right| + \left|\int_0^T \left(\overline{W}\left(f-F\right), \varphi\right)\mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(\overline{W}F, \varphi-\varphi_h\right)\mathrm{d}t\right| + \left|\int_0^T \left(\varphi_h F - R_h(\varphi_h F), \overline{W}\right)\mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(R_h(\varphi_h F), \overline{W}\right) - \left(R_h(\varphi_h F), \overline{W}\right)_h \mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(w-\overline{W}, f\varphi\right)\mathrm{d}t\right| \\
&\qquad + \left\|\overline{W}\right\|_{L^2(0,T;L^2)} \|f-F\|_{L^2(0,T;L^2)} \|\varphi\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + \left\|\overline{W}\right\|_{L^2(0,T;L^2)} \|F\|_{L^\infty(0,T;L^\infty)} \|\varphi-\varphi_h\|_{L^2(0,T;L^2)} \\
&\qquad + ch \left\|\overline{W}\right\|_{L^2(0,T;L^2)} \|\varphi_h\|_{L^2(0,T;H^1)} \|F\|_{L^\infty(0,T;L^\infty)} \\
&\qquad + ch \left\|\overline{W}\right\|_{L^2(0,T;L^2)} \|F\|_{L^2(0,T;H^1)} \|\varphi_h\|_{L^\infty(0,T;L^\infty)} \\
&\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}$$

by (3.4.8), (3.2.3), (3.4.16), (3.4.17), and the properties of the Ritz projection and the Lagrange interpolation.

Integrating by parts, the next term gives

$$\begin{aligned}
&\left|\int_0^T B_8\mathrm{d}t - \int_0^T A_8\mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(w-\overline{W}, \Delta\varphi\right)\mathrm{d}t\right| + \left|\int_0^T \left(\nabla\overline{W}, \nabla\left(\varphi_h-\varphi\right)\right)\mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(w-\overline{W}, \Delta\varphi\right)\mathrm{d}t\right| + c\left\|\nabla\overline{W}\right\|_{L^2(0,T;L^2)} \|\varphi-\varphi_h\|_{L^2(0,T;H^1)} \\
&\quad\le \left|\int_0^T \left(w-\overline{W}, \Delta\varphi\right)\mathrm{d}t\right| + ch^{1/2} \|\varphi\|_{L^2(0,T;H^2)} \\
&\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}$$

by (3.4.8), an inverse estimate, the properties of the Lagrange interpolation, and (3.4.7).

Next,

$$\begin{aligned}
&\left|\int_0^T B_9\mathrm{d}t - \int_0^T A_9\mathrm{d}t\right| \\
&\quad\le c\|\nabla s\|_{L^2(0,T;L^2)} \|\nabla\left(\psi-\psi_h\right)\|_{L^2(0,T;L^2)} + c\left\|\nabla\left(s-\overline{S}\right)\right\|_{L^2(0,T;L^2)} \|\nabla\psi_h\|_{L^2(0,T;L^2)} \\
&\xrightarrow[h\to 0]{k\to 0} 0,
\end{aligned}$$

by (3.4.8) and the properties of the Lagrange interpolation.

And finally,

$$\begin{aligned}
&\left|\int_0^T B_{10}\mathrm{d}t - \int_0^T A_{10}\mathrm{d}t\right| \\
&\quad\le \left|\int_0^T \left(\overline{W}, \psi - \psi_h\right)\mathrm{d}t\right| + \left|\int_0^T \left(w - \overline{W}, \psi\right)\mathrm{d}t\right| \\
&\qquad + \left|\int_0^T \left(\overline{W}, \psi_h\right) - \left(\overline{W}, \varphi_h\right)_h \mathrm{d}t\right| \\
&\quad\le c\left\|\overline{W}\right\|_{L^2(0,T;L^2)} \left\|\psi - \psi_h\right\|_{L^2(0,T;L^2)} + \left|\int_0^T \left(w - \overline{W}, \psi\right)\mathrm{d}t\right| \\
&\qquad + ch\left\|\overline{W}\right\|_{L^2(0,T;L^2)} \left\|\nabla\psi_h\right\|_{L^2(0,T;L^2)} \\
&\xrightarrow[h\to 0]{k\to 0} 0.
\end{aligned}$$

This concludes the proof.

□

3.5 Mumford-Shah-Euler with Lagrange Multiplier

The system (3.3.2) penalises the sphere constraint $|\mathbf{u}| = 1$ by a Ginzburg-Landau term that is scaled by $\delta \equiv \delta_\varepsilon > 0$. According to Section 3.4, weak solutions $(\mathbf{u}_\delta, s_\delta)$ of (3.3.2) may be constructed as proper limits of iterates $\{\mathbf{U}_n^\delta, S_n^\delta\}_n$ which solve the implementable space-time discretisation (3.4.2).

Passing to the limit for $\delta \to 0$ should be done simultaneously with ε in the context of Γ-convergence of the functional (3.3.1) (i.e., the related minimisation problem), which we do not study here. We would expect a moderate condition on δ converging to zero not too fast compared to ε, similar to the one in [67] and Chapter 2, compare also the computational studies in Section 3.6.

Passing to the limit for $\delta \to 0$ in equation (3.3.2), on the other hand, seems non-trivial: Although sufficient compactness properties are provided by the energy law for suitably relabelled sequences $\{(\mathbf{u}_\delta, s_\delta)\}_\delta$ to accomplish this goal in $(3.3.2)_1$, strong convergence $\nabla\mathbf{u}_\delta \to \nabla\mathbf{u}$ in $L^2(\Omega_T)$ $(\delta \to 0)$ may not be expected in general; c.f. [98, p. 283ff.]. This property, however, is needed to conclude for the second term in $(3.3.2)_2$ that $|\nabla\mathbf{u}_\delta|^2 s_\delta \rightharpoonup |\nabla\mathbf{u}|^2 s$ in $L^2(\Omega_T)$ $(\delta \to 0)$. It is this term that controls the interplay of the sphere-valued heat flow harmonic map with the phase-field evolution. We also refer to Step 7 in the proof of Theorem 3.4.2, which is crucial for proving (sub-) convergence of iterates of Algorithm 3.4.1 to weak solutions of (3.3.2) for $(k, h) \to 0$.

In this section, we propose a consistent discretisation of the Mumford-Shah-Euler flow without penalisation:

$$\begin{aligned} \mathbf{u}_t - \gamma \mathrm{div}\left(\left(s^2+k_\varepsilon\right)\nabla\mathbf{u}\right) + \lambda\chi_B\left(\mathbf{u}-\mathbf{u}_0\right) &= \mu\mathbf{u}, \\ s_t + \gamma|\nabla\mathbf{u}|^2 s + \frac{\alpha}{\varepsilon}\psi'(s) - \alpha\varepsilon w & \\ +2\beta\left(\frac{1}{\varepsilon^3}\psi''(s)\psi'(s) - \frac{1}{\varepsilon}\Delta\psi'(s) - \frac{1}{\varepsilon^2}\psi''(s)w + \Delta w\right) &= 0, \\ -\varepsilon\Delta s + w &= 0, \\ |\mathbf{u}| &= 1, \end{aligned} \tag{3.5.1}$$

with appropriate initial and boundary conditions, where $\mu : \Omega_T \to \mathbb{R}$ is the Lagrange multiplier for the sphere constraint $|\mathbf{u}| = 1$. By scalar multiplication of $(3.5.1)_1$ with $\mathbf{u}$, noting $(3.5.1)_4$, we easily compute

$$\mu = \gamma\left(s^2+k_\varepsilon\right)|\nabla\mathbf{u}|^2 + \lambda\chi_B\left(1-\mathbf{u}_0\cdot\mathbf{u}\right). \tag{3.5.2}$$

Moreover, multiplication of $(3.5.1)_1$ with $\mathbf{u}_t$, and $(3.5.1)_2$ with s_t, integration of over Ω_T, and summation formally gives

$$E_\varepsilon(\mathbf{u}(T,\cdot),s(T,\cdot)) + \|\mathbf{u}_t\|^2_{L^2(0,T;L^2)} + \|s_t\|^2_{L^2(0,T;L^2)} = E_\varepsilon(\mathbf{u}(0,\cdot),s(0,\cdot)), \tag{3.5.3}$$

where E_ε is now defined as

$$\begin{aligned} E_\varepsilon(\mathbf{u},s) :=& \frac{\gamma}{2}\int_\Omega \left(s^2+k_\varepsilon\right)|\nabla\mathbf{u}|^2\,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_B |\mathbf{u}-\mathbf{u}_0|^2\,\mathrm{d}\mathbf{x} \\ &+ \alpha\int_\Omega\left(\frac{\varepsilon}{2}|\nabla s|^2 + \frac{1}{\varepsilon}\psi(s)\right)\mathrm{d}\mathbf{x} + \frac{\beta}{\varepsilon}\int_\Omega\left(\frac{1}{\varepsilon}\psi'(s)-\varepsilon\Delta s\right)^2\mathrm{d}\mathbf{x}, \end{aligned}$$

for $\mathbf{u}\in H^1(\Omega,\mathbb{S}^{m-1})$, $\mathbf{u}_0\in L^\infty(\Omega,\mathbb{S}^{m-1})$, and $s\in H^2(\Omega)$.

3.5.1 Algorithm

We combine the results from Section 3.4 and [11] to construct an implementable space-time discretisation of (3.5.1) where the sphere constraint is preserved at each node $\mathbf{z}\in\mathcal{N}_h$, and a discrete version of (3.5.3) is valid. The discrete energy that we consider in this section is

$$\begin{aligned} E_{h,\varepsilon}(\mathbf{U},S) :=& \frac{\gamma}{2}\int_\Omega \left(S^2+k_\varepsilon\right)|\nabla\mathbf{U}|^2\,\mathrm{d}\mathbf{x} + \frac{\lambda}{2}\int_B |\mathbf{U}-\mathbf{U}_0|^2\,\mathrm{d}\mathbf{x} \\ &+ \alpha\int_\Omega\left(\frac{\varepsilon}{2}|\nabla S|^2 + \frac{1}{\varepsilon}\psi(S)\right)\mathrm{d}\mathbf{x} + \frac{\beta}{\varepsilon}\int_\Omega \mathcal{I}_h\left(\left(\frac{1}{\varepsilon}\psi'(S)-\varepsilon\widetilde{\Delta}_h S\right)^2\right)\mathrm{d}\mathbf{x}, \end{aligned}$$

for $\mathbf{U},\mathbf{U}_0\in V_h(\Omega,\mathbb{R}^m)$, and $S\in V_h(\Omega)$ (only the last term has changed from E_ε above).

Algorithm 3.5.1. *Let $\mathbf{U}_0\in V_h(\Omega,\mathbb{S}^{m-1})$, and $S_0\in V_h(\Omega)$ be given.*

(1) *Compute $W_0\in V_h(\Omega)$, such that $(W_0,Z) = -\varepsilon\left(\nabla S_0,\nabla Z\right)$ for all $Z\in V_h(\Omega)$.*

(2) *For $n = 0, \dots$ compute $S_{n+1}, W_{n+1} \in V_h(\Omega)$, and $\mathbf{U}_{n+1} \in V_h(\Omega, \mathbb{R}^m)$ such that for all $Y, Z \in V_h(\Omega)$ and $\mathbf{X} \in V_h(\Omega, \mathbb{R}^m)$ the following equations hold:*

$$\begin{aligned}
(d_t \mathbf{U}_{n+1}, \mathbf{X})_h + \frac{\gamma}{2}\left(\left(S_{n+1}^2 + S_n^2 + 2k_\varepsilon\right)\nabla \mathbf{U}_{n+1/2}, \nabla \mathbf{X}\right) \\
+\lambda\left(\chi_B\left(\mathbf{U}_{n+1/2} - \mathbf{U}_0\right), \mathbf{X}\right) - \left(\mu_{n+1}\mathbf{U}_{n+1/2}, \mathbf{X}\right)_h = 0, \\
(d_t S_{n+1}, Y) + \frac{\gamma}{2}\left(\left(|\nabla \mathbf{U}_{n+1}|^2 + |\nabla \mathbf{U}_n|^2\right) S_{n+1/2}, Y\right) \\
+\frac{2\alpha}{\varepsilon}\left(\left(S_{n+1}^2 + S_n^2 - 2\right) S_{n+1/2}, Y\right) + \alpha\varepsilon\left(\nabla S_{n+1/2}, \nabla Y\right) \\
+\frac{16\beta}{\varepsilon^3}\Big(\left(S_{n+1}^2 + S_{n+1}S_n + S_n^2 - 1\right)\Big(\left(S_{n+1}^2 - 1\right) S_{n+1} \\
+\left(S_n^2 - 1\right) S_n\Big), Y\Big)_h \\
+\frac{4\beta}{\varepsilon}\left(\nabla \mathcal{I}_h\big(\left(S_{n+1}^2 - 1\right) S_{n+1} + \left(S_n^2 - 1\right) S_n\big), \nabla Y\right) \\
-\frac{8\beta}{\varepsilon^2}\left(\left(S_{n+1}^2 + S_{n+1}S_n + S_n^2 - 1\right) W_{n+1/2}, Y\right)_h - 2\beta\left(\nabla W_{n+1/2}, \nabla Y\right) = 0, \\
\varepsilon\left(\nabla S_{n+1}, \nabla Z\right) + \left(W_{n+1}, Z\right)_h = 0.
\end{aligned} \tag{3.5.4}$$

Here, $\mu_{n+1} \in V_h(\Omega)$ is the approximate discrete Lagrange multiplier to establish the discrete sphere constraint, for which we also have to use reduced integration in the first term. An an explicit formula for μ_{n+1} is (for $\mathbf{z} \in \mathcal{N}_h$)

$$\mu_{n+1}(\mathbf{z}) = \begin{cases} 0 & \text{if } \mathbf{U}_{n+1/2}(\mathbf{z}) = 0, \\ \frac{\frac{\gamma}{2}\left(\left(S_{n+1}^2 + S_n^2 + 2k_\varepsilon\right)\nabla \mathbf{U}_{n+1/2}, \mathbf{U}_{n+1/2}(\mathbf{z}) \otimes \nabla \varphi_{\mathbf{z}}\right)}{\beta_{\mathbf{z}}\left|\mathbf{U}_{n+1/2}(\mathbf{z})\right|^2} & \\ \quad + \frac{\lambda \chi_B(\mathbf{z})\left(\left|\mathbf{U}_{n+1/2}(\mathbf{z})\right|^2 - \mathbf{U}_0(\mathbf{z}) \cdot \mathbf{U}_{n+1/2}(\mathbf{z})\right)}{\left|\mathbf{U}_{n+1/2}(\mathbf{z})\right|^2} & \text{otherwise.} \end{cases} \tag{3.5.5}$$

This formula can be easily found using $\mathbf{X} = \mathbf{U}_{n+1/2}(\mathbf{z})\varphi_{\mathbf{z}}$, with the properties of $(\cdot, \cdot)_h$. We remark that the non-local character of (3.5.5) as opposed to its continuous analog (3.5.2) is due to space-time discretisation.

3.5.2 Analysis

The following proposition asserts existence of solutions $\{(\mathbf{U}_{n+1}, S_{n+1})\}_n$ of $(3.5.4)_{1-3}$, (3.5.5), which avoids a mixed formulation for quantities $\{(\mathbf{U}_{n+1}, \mu_{n+1}, S_{n+1})\}_n$ in Algorithm 3.5.1 for every $n \geq 0$. This reduction of the original problem requires a mesh constraint to hold, to exclude the case $\mathbf{U}_{n+1/2}(\mathbf{z}) = 0$ in (3.5.5).

Proposition 3.5.2. *Let $\mathcal{T}_h$ be a quasi-uniform triangulation of $\Omega \subset \mathbb{R}^d$, $\mathbf{U}^0 \in V_h(\Omega, \mathbb{R}^m)$ such that $|\mathbf{U}_0(\mathbf{z})| = 1$ for all $\mathbf{z} \in \mathcal{N}_h$, and $d \leq 3$. For sufficiently small $\widetilde{C} \equiv \widetilde{C}(\Omega, \mathcal{T}_h) > 0$ independent of $k, h > 0$ such that $k \leq \widetilde{C} \max\left\{h^{d+2}, \varepsilon^{-1}h^4\right\}$, there exists a tuple $\{\mathbf{U}_{n+1}, S_{n+1}\}_{n \geq 0}^N \subset V_h(\Omega, \mathbb{R}^m) \times V_h(\Omega)$ that satisfies $(3.5.4)_{1-3}$, (3.5.5), $|\mathbf{U}_{n+1}(\mathbf{z})| = 1$ for all $\mathbf{z} \in \mathcal{N}_h$, and $(1 \leq n \leq N)$*

$$E_{h,\varepsilon}\left(\mathbf{U}^{n+1}, S^{n+1}\right) + k\sum_{\ell=0}^{n}\left(\left\|d_t \mathbf{U}_{\ell+1}\right\|_{L^2}^2 + \left\|d_t S^{\ell+1}\right\|_{L^2}^2\right) = E_{h,\varepsilon}(\mathbf{U}_0, S_0)\,.$$

In passing we remark that for $m = 3$, the heat flow harmonic map to $\mathbb{S}^2$ which weakly solves $\mathbf{v}_t - \Delta\mathbf{v} = |\nabla\mathbf{v}|^2\mathbf{v}$ a.e. in Ω_T also satisfies $\mathbf{v}_t + \mathbf{v} \times (\mathbf{v} \times \Delta\mathbf{v}) = 0$ in Ω_T.

This cross-product formulation was studied in [12] to get another convergent discretisation of the heat flow harmonic map equation to the sphere $\mathbb{S}^2$. But such a strategy is not useful in our case, where a discretisation of (3.5.1) that satisfies a discrete energy law requires a proper balancing of corresponding terms in $(3.5.1)_1$ and $(3.5.1)_2$: In order to obtain the energy law for the cross-product formulation, we have to multiply by $-\Delta\mathbf{u}$ rather than $\mathbf{u}_t$ before integration in space and time,

$$\frac{1}{2}\left\|\nabla\mathbf{u}(T,\cdot)\right\|^2 + \int_{\Omega_T} \left|\mathbf{u} \times \Delta\mathbf{u}\right|^2 \mathrm{d}\mathbf{x}\mathrm{d}t = \frac{1}{2}\left\|\nabla\mathbf{u}(0,\cdot)\right\|^2.$$

Hence, a corresponding reformulation of $(3.5.1)_1$ requires a test function different from $\mathbf{u}_t$, which conflicts with the necessary choice of s_t in $(3.5.1)_2$ to eventually establish an energy law for (3.5.1).

The following proof per induction of Proposition 3.5.2 adopts arguments from [11] to $(3.5.4)_{1-3}$, (3.5.5):

In a first step, we modify the discrete Lagrange multiplier in (3.5.5) to exclude 'division by zero', control the interplay of phase field and harmonic map evolution in the coupling term, and obtain existence of corresponding solutions by Brouwer's fixed point theorem.

A bootstrapping argument afterwards validates that $\mathbf{U}_{n+1/2}(\mathbf{z}) = 0$ for any $\mathbf{z} \in \mathcal{N}_h$ is excluded once $k \leq \widetilde{C}h^{d+2}$ is valid, so in this case, the mesh constraint stems from the discrete Lagrange multiplier.

Proof. **Step 1:** *Solvability of an auxiliary problem.*

Fix $n \geq 0$. For every $0 < \kappa \leq \frac{1}{4}$, define the continuous mapping

$$\mathbf{F}_\kappa : V_h(\Omega, \mathbb{R}^m) \times V_h(\Omega) \times V_h(\Omega) \to V_h(\Omega, \mathbb{R}^m) \times V_h(\Omega) \times V_h(\Omega)$$

via

$$
\begin{aligned}
&(\mathbf{F}_\kappa(\mathbf{\Phi},\Theta,\Xi),(\mathbf{X},Y,Z))\\
&:= \frac{2}{k}\left(\mathbf{\Phi}-\mathbf{U}_n,\mathbf{X}\right)_h + \frac{\gamma}{2}\left(\left((2\Theta-S_n)^2+S_n^2+2k_\varepsilon\right)\nabla\mathbf{\Phi},\nabla\mathbf{X}\right)\\
&\quad-\sum_{\mathbf{z}\in\mathcal{N}_h}\left(\lambda\left(\chi_B(\mathbf{z})\frac{|\mathbf{\Phi}(\mathbf{z})|^2-\mathbf{U}_0(\mathbf{z})\cdot\mathbf{\Phi}(\mathbf{z})}{\max\left\{|\mathbf{\Phi}(\mathbf{z})|^2,\kappa\right\}}\varphi_{\mathbf{z}}\mathbf{\Phi},\mathbf{X}\right)_h\right.\\
&\quad\left.+\frac{\gamma}{2}\left(\frac{\left(\left(\min\left\{(2\Theta-S_n)^2,\kappa^{-1}\right\}+S_n^2+k_\varepsilon\right)\nabla\mathbf{\Phi},\mathbf{\Phi}(\mathbf{z})\otimes\nabla\varphi_{\mathbf{z}}\right)}{\beta_{\mathbf{z}}\max\left\{|\mathbf{\Phi}(\mathbf{z})|^2,\kappa\right\}}\varphi_{\mathbf{z}}\mathbf{\Phi},\mathbf{X}\right)_h\right)\\
&\quad+\frac{2}{k}\left(\Theta-S_n,Y\right)+\frac{\gamma}{2}\left(\left(|\nabla\left(2\mathbf{\Phi}-\mathbf{U}_n\right)|^2+|\nabla\mathbf{U}_n|^2\right)\Theta,Y\right)\\
&\quad+\frac{2\alpha}{\varepsilon}\left(\left((2\Theta-S_n)^2+S_n^2-2\right)\Theta,Y\right)+\alpha\varepsilon\left(\nabla\Theta,\nabla Y\right)\\
&\quad-\frac{16\beta}{\varepsilon^3}\left(\left((2\Theta-S_n)^2+2\Theta\cdot S_n-1\right)\left(\left(1-(2\Theta-S_n)^2\right)(2\Theta-S_n)+\left(1-S_n^2\right)S_n\right),Y\right)_h\\
&\quad+\frac{4\beta}{\varepsilon}\left(\nabla\mathcal{I}_h\left(\left((2\Theta-S_n)^2-1\right)(2\Theta-S_n)+\left(S_n^2-1\right)S_n\right),\nabla Y\right)\\
&\quad-\frac{8\beta}{\varepsilon^2}\left(\left((2\Theta-S_n)^2+(2\Theta-S_n)S_n+S_n^2-1\right)\Xi,Y\right)_h-2\beta\left(\nabla\Xi,\nabla Y\right)\\
&\quad+2\beta(\nabla\Theta,\nabla Z)+2\frac{\beta}{\varepsilon}(\Xi,Z)_h.
\end{aligned}
$$

Setting $(\mathbf{X},Y,Z)=(\mathbf{\Phi},\Theta,\Xi)$ leads to terms $I_1,\dots I_{14}$, where for some finite $\varrho>0$

$$\sum_{i=5}^{14} I_i \geq 0 \qquad \forall\,\Theta\in V_h : \|\Theta\|_{L^\infty}\geq\varrho$$

as in Step 4 in the proof of Theorem 3.4.2.

An upper bound for $I_3:=\sum_{\mathbf{z}\in\mathcal{N}_h} I_3(\mathbf{z})$ uses $|\mathbf{U}_0(\mathbf{z})|=1$ to conclude that

$$
\begin{aligned}
|I_3(\mathbf{z})| &\leq \lambda\chi_B(\mathbf{z})\beta_{\mathbf{z}}\frac{|\mathbf{\Phi}(\mathbf{z})|^3\left(|\mathbf{\Phi}(\mathbf{z})|+|\mathbf{U}_0(\mathbf{z})|\right)}{\max\left\{|\mathbf{\Phi}(\mathbf{z})|^2,\kappa\right\}}\\
&\leq C\chi_B(\mathbf{z})\beta_{\mathbf{z}}\left|\mathbf{\Phi}(\mathbf{z})\right|\left(1+|\mathbf{\Phi}(\mathbf{z})|\right),
\end{aligned}
$$

and hence $I_3\leq C\left\|\mathbf{\Phi}\right\|_h^2$.

We compute for $I_4:=\sum_{\mathbf{z}\in\mathcal{N}_h} I_4(\mathbf{z})$,

$$
\begin{aligned}
&|I_4(\mathbf{z})|\\
&\quad\leq\frac{\gamma}{2}\frac{\left(\min\left\{(2\Theta(\mathbf{z})-S_n(\mathbf{z}))^2,\kappa^{-1}\right\}+\|S_n\|_{L^\infty}^2+2k_\varepsilon\right)|\mathbf{\Phi}(\mathbf{z})|^2}{\max\left\{|\mathbf{\Phi}(\mathbf{z})|^2,\kappa\right\}}\left(|\nabla\mathbf{\Phi}|,|\mathbf{\Phi}(\mathbf{z})\otimes\nabla\varphi_{\mathbf{z}}|\right)\\
&\quad\leq C\gamma\left(\min\left\{(2\Theta(\mathbf{z})-S_n(\mathbf{z}))^2,\kappa^{-1}\right\}+\|S_n\|_{L^\infty}^2+2k_\varepsilon\right)\left(|\nabla\mathbf{\Phi}|,|\nabla\left(\mathbf{\Phi}(\mathbf{z})\varphi_{\mathbf{z}}\right)|\right)_{\mathrm{supp}(\nabla\varphi_{\mathbf{z}})}\\
&\quad\leq Ch^{-1}\left(\kappa^{-1}+\|S_n\|_{L^\infty}^2+2k_\varepsilon\right)\left(\gamma\left|\nabla\mathbf{\Phi}\right|,|\mathbf{\Phi}(\mathbf{z})|\right)_{\mathrm{supp}(\nabla\varphi_{\mathbf{z}})}.
\end{aligned}
$$

For all $\mathbf{\Phi} \in V_h\big(\Omega, \mathbb{R}^3\big)$ such that

$$\|\mathbf{\Phi}\|_h \geq \varrho_1 := \max\left\{\|\mathbf{U}_n\|_h, \|S_n\|_{L^\infty}^4, \kappa^{-2}\right\},$$

and values $k \leq \widetilde{C}h^2$ for some existing $0 < \widetilde{C} \equiv \widetilde{C}(\Omega)$, on using Young's inequality, and the fact that the number of nodes $\mathbf{z} \in \mathcal{N}_h$ such that $(\nabla\varphi_{\mathbf{y}}, \nabla\varphi_{\mathbf{z}}) \neq 0$ is bounded independently of $h > 0$,

$$\begin{aligned} I_1 + \ldots + I_4 \;&\geq\; \frac{2}{k}\left(\mathbf{\Phi} - \mathbf{U}_n, \mathbf{\Phi}\right)_h + \frac{\gamma}{2}\left\|\nabla\mathbf{\Phi}\sqrt{(2\Theta - S_n)^2 + S_n^2 + 2k_\varepsilon}\right\|_{L^2}^2 \\ &\quad -C\|\mathbf{\Phi}\|_h^2 - Ch^{-2}k_\varepsilon^{-1}\left(\kappa^{-2} + \|S_n\|_{L^\infty}^4\right)\|\mathbf{\Phi}\|_{L^2}^2 - \gamma k_\varepsilon\|\nabla\mathbf{\Phi}\|_{L^2}^2 \\ &\geq\; \frac{2}{k}\|\mathbf{\Phi}\|_h\left(\left(1 - \frac{Ck}{k_\varepsilon\kappa^2h^2}\right)\|\mathbf{\Phi}\|_h - \|\mathbf{U}_n\|_h\right), \end{aligned}$$

Putting things together gives $\left(\mathbf{F}_\kappa(\mathbf{\Phi}, \Theta, \Xi), (\mathbf{\Phi}, \Theta, \Xi)\right) \geq 0$ for all $(\mathbf{\Phi}, \Theta, \Xi)$ sufficiently large. A corollary to Brouwer's fixed-point theorem [65, p. 279] then implies existence of a tuple $\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big) \in V_h(\Omega, \mathbb{R}^m) \times V_h(\Omega) \times V_h(\Omega)$ such that $\mathbf{F}_\kappa\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big) = 0$.

Step 2: *Show that* $\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big)$ *solves* (3.5.4) *for* $k \leq \widetilde{C}h^{d+2}$.

We proceed by induction to show that the root $\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big)$ of $\mathbf{F}_\kappa$ solves $\mathbf{F}_0\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big) = 0$, if $k \leq \widetilde{C}h^{d+2}$. Let $n \geq 1$. For all $0 \leq \ell \leq n$, suppose that $\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big) \in V_h(\Omega, \mathbb{R}^m) \times V_h(\Omega) \times V_h(\Omega)$ satisfies

$$\begin{aligned} |\mathbf{U}_\ell(\mathbf{z})| &= 1 \qquad \forall\, \mathbf{z} \in \mathcal{N}_h, \\ E_{h,\varepsilon}(\mathbf{U}_\ell, S_\ell) + \frac{k}{2}\sum_{j=0}^{\ell-1}\left(\|d_t\mathbf{U}_j\|_h^2 + \|d_tS_j\|_{L^2}^2\right) &= E_{h,\varepsilon}(\mathbf{U}_0, S_0)\,. \end{aligned} \tag{3.5.6}$$

We now want to show that, if $k \leq \widetilde{C}h^{d+2}$ for some $\widetilde{C} \geq 0$, then

$$\begin{aligned} &\|S_{n+1}\|_{L^\infty(\Omega)} \leq h^{-d/2}C, \text{ for some } C \geq 0 \text{ independent of } n, \text{ and} \\ &\left|\mathbf{U}_{n+1/2}(\mathbf{z})\right| > 0. \end{aligned} \tag{3.5.7}$$

For $(3.5.7)_1$, we use arguments similar to the ones in the proof of Theorem 3.4.2, Step 4 (the energy principle): From Step 1 above we know that $\mathbf{F}_\kappa\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big) = 0$. Now consider

$$\begin{aligned} 0 &= \left(\mathbf{F}_\kappa\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big), \big(\mathbf{0}, S_{n+1/2}, W_{n+1/2}\big)\right) \\ &=: T_1 + \ldots + T_{14}. \end{aligned}$$

We can assume that

$$\|S_{n+1}\|_{L^2(\Omega)} \geq \varrho_2\Big(\|S_n\|_{L^2(\Omega)}, |\Omega|\Big),$$

since otherwise, by inverse estimate and $(3.5.6)_2$,

$$\|S_{n+1}\|_{L^\infty(\Omega)} \leq Ch^{-d/2}\|S_{n+1}\|_{L^2(\Omega)} \leq Ch^{-d/2}\|S_n\|_{L^2(\Omega)} \leq Ch^{-d/2},$$

and we would be done.

In the following, we repeatedly use Young's and Hölder's inequalities, and standard inverse estimates, see e.g. [24, Section 4.5]. Furthermore, we use the fact that by (3.5.6), $\|\nabla \mathbf{U}_n\|_{L^2}^2 \leq c_0 k_\varepsilon^{-1} = c(E_{h,\varepsilon}(\mathbf{U}_0, S_0))\, k_\varepsilon^{-1}$, $\|\nabla S_n\|_{L^2}^2 \leq c_0 \varepsilon^{-1}$, as well as $\|S_n\|_{L^\infty} \leq c_0 \varepsilon^{-5/2}$, $\|\nabla S_n\|_{L^6} \leq c_0 \varepsilon^{-4}$, and $\|W_n\|_{L^2}^2 \leq c_0 \varepsilon$.

By the choice of the test function, $\sum_{i=1}^4 T_i = 0$, and

$$T_5 = \frac{1}{k} \|S_{n+1}\|_{L^2(\Omega)}^2 - \frac{1}{k} \|S_n\|_{L^2(\Omega)}^2 .$$

Like in the proof of Theorem 3.4.2, Step 4,

$$T_6 \;=\; \frac{\gamma}{2} \left(|\nabla \mathbf{U}_{n+1}|^2 + |\nabla \mathbf{U}_n|^2, \left|S_{n+1/2}\right|^2 \right) \geq 0,$$

$$T_7 = \frac{2\alpha}{\varepsilon} \left(\left(S_{n+1}^2 + S_n^2 - 2\right), \left|S_{n+1/2}\right|^2 \right) \geq 0,$$

if $\|S_{n+1}\|_{L^2} \geq \varrho_2\Big(\|S_n\|_{L^2(\Omega)}, |\Omega| \Big)$; and

$$T_8 = \alpha\varepsilon \left\|\nabla S_{n+1/2}\right\|_{L^2(\Omega)}^2 \geq 0.$$

By the same calculations as in the proof of Theorem 3.4.2, Step 4,

$$\begin{aligned} &\frac{\varepsilon}{2\beta} \left(T_9 + T_{10} + T_{11} + T_{12} + T_{13} + T_{14}\right) \\ &\quad \geq \frac{1}{\varepsilon^2} \left\| \left|2S_{n+1/2}\right|^3 \right\|_h^2 - \frac{c\varepsilon^2}{h^4} \left\|S_{n+1/2}\right\|_{L^2(\Omega)}^2 , \end{aligned}$$

and this remainder can be compensated by T_5, by the coupling between k and h.

Therefore, by an inverse estimate,

$$\|S_{n+1}\|_{L^\infty(\Omega)} \leq C h^{-d/2} \|S_{n+1}\|_{L^2(\Omega)} \leq C h^{-d/2} \|S_n\|_{L^2(\Omega)} \leq C h^{-d/2}.$$

For $(3.5.7)_2$, by parallelogram identity and triangle inequality, it suffices to show

$$\left|\mathbf{U}_{n+1/2}(\mathbf{z})\right|^2 \geq 1 - \frac{k}{2} \left|d_t \mathbf{U}_{n+1}(\mathbf{z})\right|^2 \overset{!}{>} \frac{1}{2} \qquad \forall\, \mathbf{z} \in \mathcal{N}_h. \tag{3.5.8}$$

By the definition of $\mathbf{F}_\kappa$, the iterate $\mathbf{U}_{n+1} = 2\mathbf{U}_{n+1/2} - \mathbf{U}_n$ satisfies for all $\mathbf{X} \in V_h(\Omega, \mathbb{R}^m)$,

$$\begin{aligned} &\left(d_t \mathbf{U}_{n+1}, \mathbf{X}\right)_h + \frac{\gamma}{2} \left(\left(S_{n+1}^2 + S_n^2 + 2k_\varepsilon\right) \nabla \mathbf{U}_{n+1/2}, \nabla \mathbf{X} \right) \\ &= \lambda \sum_{\mathbf{z} \in \mathcal{N}_h} \left(\chi_B(\mathbf{z}) \frac{\left|\mathbf{U}_{n+1/2}(\mathbf{z})\right|^2 - \mathbf{U}_0(\mathbf{z}) \cdot \mathbf{U}_{n+1/2}(\mathbf{z})}{\max\{\left|\mathbf{U}_{n+1/2}(\mathbf{z})\right|^2, \kappa\}} \varphi_{\mathbf{z}} \mathbf{U}_{n+1/2}, \mathbf{X} \right)_h \\ &\quad + \sum_{\mathbf{z} \in \mathcal{N}_h} \left(\frac{\gamma \left(\left(\min\left\{S_{n+1}^2, \kappa^{-1}\right\} + S_n^2 + 2k_\varepsilon\right) \nabla \mathbf{U}^{n+1/2}, \mathbf{U}^{n+1/2}(\mathbf{z}) \otimes \nabla \varphi_{\mathbf{z}} \right)}{2\beta_{\mathbf{z}} \max\{|\mathbf{U}^{n+1/2}(\mathbf{z})|^2, \kappa\}} \varphi_{\mathbf{z}} \mathbf{U}^{n+1/2}, \mathbf{X} \right)_h . \end{aligned}$$

We put $\mathbf{X} = d_t\mathbf{U}_{n+1}(\mathbf{z}^*)\,\varphi_{\mathbf{z}^*}$ for $\mathbf{z}^* = \operatorname{argmax}_{\mathbf{y}\in\mathcal{N}_h} |d_t\mathbf{U}^{n+1}(\mathbf{y})|$, and use properties of reduced integration, an inverse estimate, inequality $(3.5.7)_1$, and $\|\nabla\varphi_{\mathbf{z}^*}\|_{L^1} \leq C\beta_{\mathbf{z}^*}h^{-1}$, to obtain

$$\begin{aligned}
&\beta_{\mathbf{z}^*}\,|d_t\mathbf{U}_{n+1}(\mathbf{z}^*)|^2\\
&\quad\leq c\,\big|\big(\big(S_{n+1}^2+S_n^2+2k_\varepsilon\big)\,\nabla\mathbf{U}_{n+1/2}, d_t\mathbf{U}_{n+1}(\mathbf{z}^*)\,\nabla\varphi_{\mathbf{z}^*}\big)\big|\\
&\qquad+c\,\big(\big|\mathbf{U}_{n+1/2}(\mathbf{z})\big|+|\mathbf{U}_0(\mathbf{z})|\big)\,\|\varphi_{\mathbf{z}^*}\|_{L^1}\,|d_t\mathbf{U}_{n+1}(\mathbf{z}^*)|\\
&\qquad+c\,\big\|S_{n+1}^2+S_n^2+2k_\varepsilon\big\|_{L^\infty}\,\big\|\nabla\mathbf{U}_{n+1/2}\big\|_{L^\infty}\,\|\nabla\varphi_{\mathbf{z}^*}\|_{L^1}\,|d_t\mathbf{U}_{n+1}(\mathbf{z}^*)|\\
&\quad\leq ch^{-d}\left(\frac{k}{2}\,\|\nabla d_t\mathbf{U}_{n+1}\|_{L^\infty}+\|\nabla\mathbf{U}_n\|_{L^\infty}\right)\|\nabla\varphi_{\mathbf{z}^*}\|_{L^1}\,|d_t\mathbf{U}_{n+1}(\mathbf{z}^*)|\\
&\quad\leq c\beta_{\mathbf{z}^*}h^{-(d+2)}\left(\frac{k}{2}\,|d_t\mathbf{U}_{n+1}(\mathbf{z}^*)|+\|\mathbf{U}_n\|_{L^\infty}\right)|d_t\mathbf{U}_{n+1}(\mathbf{z}^*)|\,.
\end{aligned}$$

By assumption (3.5.6) for all $0\leq\ell\leq n$, we then arrive at

$$\left(1-\frac{ck}{h^{d+2}}\right)|d_t\mathbf{U}_{n+1}(\mathbf{z}^*)|\leq\tilde{c}h^{-(d+2)},$$

for some $\tilde{c}\equiv\tilde{c}(\Omega)>0$. Hence, assertion (3.5.8) is valid for values $k\leq\tilde{c}h^{d+2}$. Consequently, the parameter $\kappa>0$ which was needed a priori to exclude division by zero and unboundedness of iterates $\|S_{n+1}\|_{L^\infty(\Omega)}$ in the Lagrange multiplier, really has no effect for k small enough; i.e., $\mathbf{F}_0\big(\mathbf{U}_{n+1/2}, S_{n+1/2}, W_{n+1/2}\big)=0$. Therefore, testing (3.5.1) with $\mathbf{X}=\mathbf{U}_{n+1/2}\varphi_{\mathbf{z}}$ and using the definition of μ_{n+1}, we verify that $|\mathbf{U}_{n+1}(\mathbf{z})|=1$ for all $\mathbf{z}\in\mathcal{N}_h$.

Moreover, the energy bound $(3.5.6)_2$ holds for all $0\leq\ell\leq n+1$, which follows like in the proof of Theorem 3.4.2, Step 2: Test (3.5.4) with $\mathbf{X}:=d_t\mathbf{U}_{n+1}$, $Y:=d_tS_{n+1}$, and observe that that $\big(\mathbf{U}_{n+1/2}, d_t\mathbf{U}_{n+1}\big)_h=\frac{1}{2k}\sum_{\mathbf{z}}\beta_{\mathbf{z}}\big(|\mathbf{U}_{n+1}(\mathbf{z})|^2-|\mathbf{U}_n(\mathbf{z})|^2\big)=0$, by $(3.5.6)_1$. □

3.6 Computational Studies

3.6.1 Implementation of Algorithms 3.4.1 and 3.5.1

For both algorithms, we use a simple fixed-point iteration to deal with the nonlinearities, and a dynamic time-stepping strategy, adjusted to the number of fixed point iterations, which are in turn adjusted to changes in the highest-order term of the energy, similarly to [53]. In order to process the brightness component, we introduce the variable $V\in V_h(\Omega)$. So, the actual implementation of Algorithm 3.4.1 looks like this:

Algorithm 3.6.1. *Let* $\mathbf{U}_0\in V_h(\Omega,\mathbb{R}^m)$, $S_0, V_0\in V_h(\Omega)$, *and* TOL >0 *be given. Set* $n:=0$ *and* $T:=0$.
Compute $W_0\in V_h(\Omega)$, *such that* $(W_0,Z)_h=-\varepsilon\,(\nabla S_0,\nabla Z)$ *for all* $Z\in V_h(\Omega)$.
While $T<1$ *do*

(1) *Set* $\mathbf{U}_{n,0}:=\mathbf{U}_{n-1}$, $V_{n,0}:=V_{n-1}$, $S_{n,0}:=S_{n-1}$, $W_{n,0}:=W_{n-1}$, *and* $\theta:=1$.

(2) *While* $\theta >$ TOL AND $l < 5$*:*

(a) Let $l := l + 1$.

(b) Compute $V_{n,l+1} \in V_h(\Omega)$*, such that for all* $X \in V_h(\Omega)$

$$\begin{aligned}(d_t V_{n+1}, X) + \frac{\gamma_1}{4}\left(\left(S_{n,l+1}^2 + S_n^2 + 2k_\varepsilon\right)\nabla\left(V_{n,l+1} + V_n\right), \nabla X\right)\\ + \frac{\lambda_1}{2}\left(\chi_B\left(V_{n,l+1} + V_n - 2V_0\right), X\right) = 0.\end{aligned}$$

(c) Compute $\mathbf{U}_{n,l+1} \in V_h(\Omega, \mathbb{R}^m)$*, such that for all* $\mathbf{X} \in V_h(\Omega, \mathbb{R}^m)$

$$\begin{aligned}(d_t \mathbf{U}_{n,l+1}, \mathbf{X}) + \frac{\gamma}{4}\left(\left(S_{n,l+1}^2 + S_n^2 + 2k_\varepsilon\right)\nabla\left(\mathbf{U}_{n,l+1} + \mathbf{U}_n\right), \nabla \mathbf{X}\right)\\ + \frac{\lambda}{2}\left(\chi_B\left(\mathbf{U}_{n,l+1} + \mathbf{U}_n - 2\mathbf{U}_0\right), \mathbf{X}\right)\\ + \frac{1}{4\delta_\varepsilon}\left(\left(|\mathbf{U}_{n,l+1}|^2 + |\mathbf{U}_n|^2 - 2\right)\left(\mathbf{U}_{n,l+1} + \mathbf{U}_n\right), \mathbf{X}\right) = 0.\end{aligned}$$

(d) Compute $S_{n,l+1}, W_{n,l+1} \in V_h(\Omega)$*, such that for all* $Y, Z \in V_h(\Omega)$

$$\begin{aligned}(d_t S_{n+1}, Y) + \frac{\gamma}{4}\left(\left(|\nabla \mathbf{U}_{n,l+1}|^2 + |\nabla \mathbf{U}_n|^2\right)\left(S_{n,l+1} + S_n\right), Y\right)\\ + \frac{\alpha}{\varepsilon}\left(\left(S_{n,l+1}^2 + S_n^2 - 2\right)\left(S_{n,l+1} + S_n\right), Y\right) + \frac{\alpha\varepsilon}{2}\left(\nabla\left(S_{n,l+1} + S_n\right), \nabla Y\right)\\ + \frac{16\beta}{\varepsilon^3}\Big(\left(S_{n,l+1}^2 + S_{n,l+1}S_n + S_n^2 - 1\right)\Big(\left(S_{n,l+1}^2 - 1\right)S_{n,l+1}\\ + \left(S_n^2 - 1\right)S_n\Big), Y\Big)_h\\ + \frac{4\beta}{\varepsilon}\left(\nabla \mathcal{I}_h\left(\left(S_{n,l+1}^2 - 1\right)S_{n,l+1} + \left(S_n^2 - 1\right)S_n\right), \nabla Y\right)\\ - \frac{4\beta}{\varepsilon^2}\left(\left(S_{n,l+1}^2 + S_{n,l+1}S_n + S_n^2 - 1\right)\left(W_{n,l+1} + W_n\right), Y\right)_h\\ - \beta\left(\nabla\left(W_{n,l+1} + W_n\right), \nabla Y\right) = 0,\\ \varepsilon\left(\nabla\left(S_{n,l+1} + S_n\right), \nabla Z\right) + \left(W_{n,l+1} + W_n, Z\right)_h = 0.\end{aligned}$$

(3) *Let* $\theta := \|W_{n,l+1} - W_{n,l}\|_{L^2(\Omega)}^2$.

(4) *If* $\theta >$ TOL*, let* $k := \frac{k}{2}$ *and go to* (2).

(5) *Else if* $l < 4$*, let* $k := 2k$.

(6) *Let* $n := n + 1$, $V_n := V_{n,l+1}$, $\mathbf{U}_n := \mathbf{U}_{n,l+1}$, $S_n := S_{n,l+1}$, $W_n := W_{n,l+1}$, *and* $T := T + k$.

Algorithm 3.5.1 is implemented in the same fashion. Since experiments show that the sphere-constraint usually improves beyond the first five fixed-point iterations, we change the first line of (2) to

While $\theta + \sigma/10 >$ TOL AND $l < 10$,
where σ is the Euclidian norm over all nodes of $1 - |\mathbf{U}_{n,l+1}|$. However, since dynamics are still primarily driven by the highest-order term, the time-step is still controlled by the number of fixed-point iterations necessary to reach $\|W_{n,l+1} - W_{n,l}\|^2_{L^2(\Omega)} <$ TOL.

All parameters in the following simulations are chosen by experiment. The parameter ε, which corresponds to the "width" of the interfaces, can usually be set to several diameters of discretisation triangles, with larger values making the interfaces more diffuse, but the dynamics more forgiving. In our simulations, we usually took $\varepsilon = 10h$, where what we call h is in fact the length of the two shorter sides of the rectangular triangles in our triangulations; i.e., it is shorter than the actual diameter of the triangles (by a factor of $\sqrt{2}$).

3.6.2 Inpainting with Penalisation, with and without Curvature

This example compares inpainting with Algorithm 3.4.1 (penalisation) with $\beta = 0$ (no curvature penalisation) versus $\beta > 0$.

Example 3.1. *Let* $\Omega := (0,1)^2$ *and* $\mathbf{U}_0, V_0$ *as in the right plot in Figure 3.1, and* $S_0 \equiv 1$. *The inpainting domain is filled with random data. Let* $h = 1/100$, $\varepsilon = 10 * h$, *and* $k_\varepsilon = \varepsilon * 10^{-3}$. *We compare the following sets of parameters:*

(1) *Let* $\gamma = \gamma_1 = 100$, $\alpha = 15$, $\beta = 0.2$, $\lambda = \lambda_1 = 5 * 10^5$, $\delta_\varepsilon = 0.05$, *and* $k_0 = h^3$ *(the initial time step; the following steps are chosen dynamically according to Section 3.6.1), and* TOL $= 10^{-2}$.

(2) *Let* $\gamma = 150$, $\gamma_1 = 130$, $\alpha = 10$, $\beta = 0$, $\lambda = \lambda_1 = 10^6$, $\delta_\varepsilon = 0.1$, *and* $k = 10^{-4}$ *(constant time-step, always 3 fixed-point iterations).*

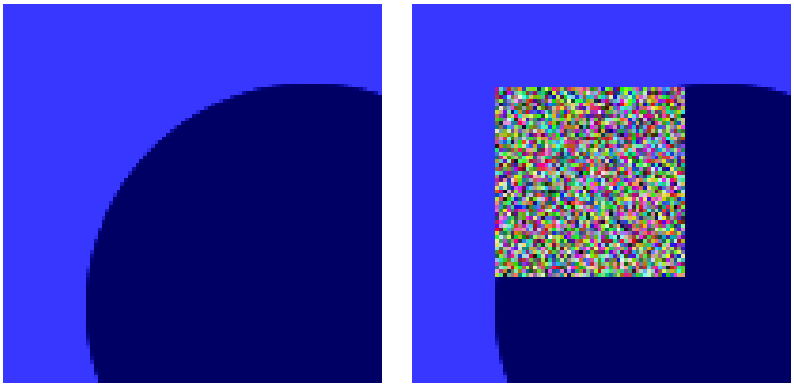

Figure 3.1: Example 3.1: Image; intact and degraded.

This example clearly shows that in some cases a curvature term is necessary to give natural inpainting results.

Figure 3.1 shows the intact image as well as the image we use as a starting point for inpainting. Figure 3.2 shows iterates, edge sets (black means $S = 0$), phase functions

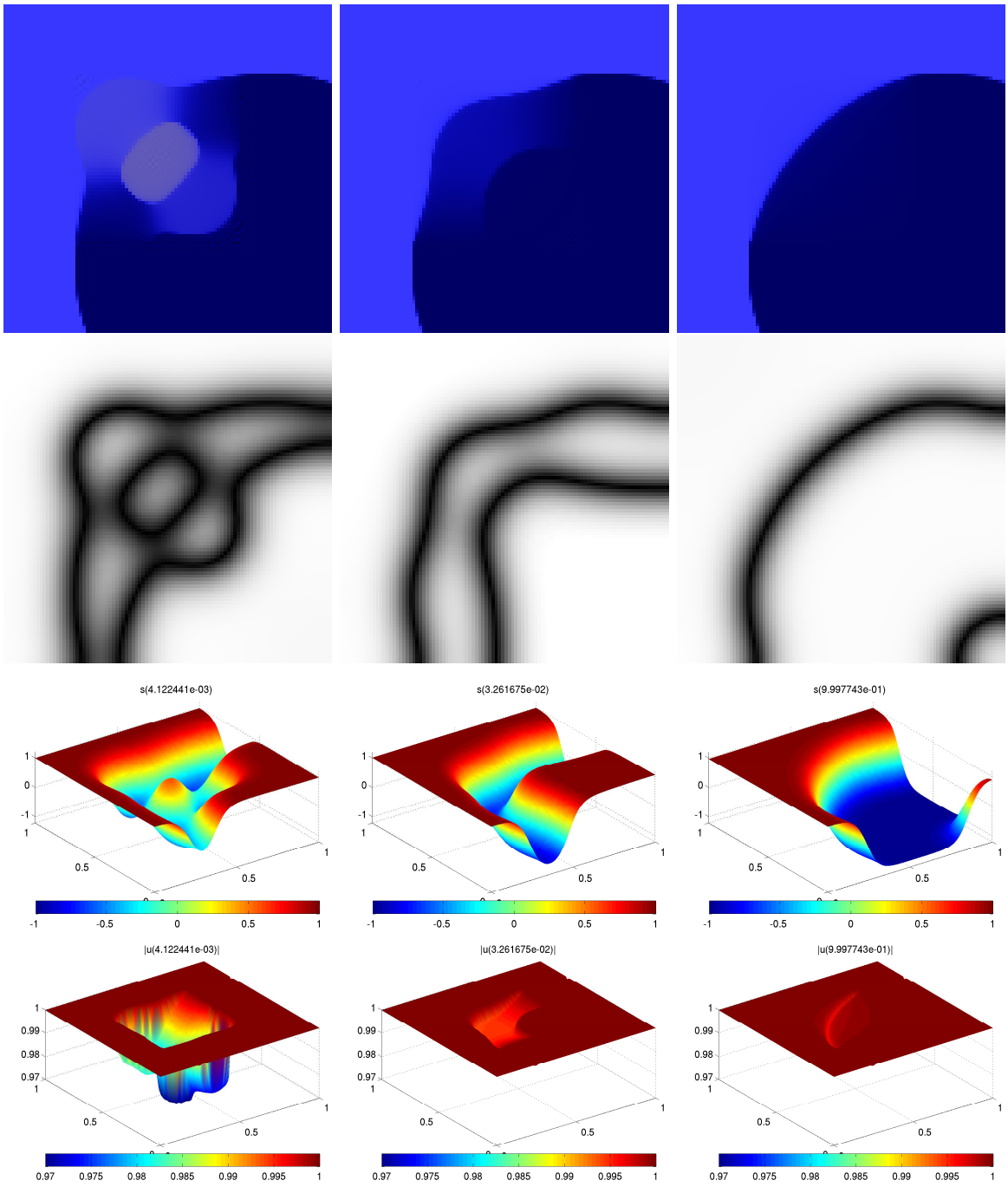

Figure 3.2: Example 3.1: $\beta = 0.2$; image (row 1), $|S|$ (row 2), S (row 3), and $|\mathbf{U}|$ (row 4) at times $t = 0.004, 0.03, 1$.

S, and the length of $\mathbf{U}$ at times $t = 0.004, 0.03, 1$ for inpainting with curvature, while Figure 3.3 shows the same at times $t = 0.009, 0.015, 1$ for inpainting without curvature term. For inpainting with curvature term we see that mainly two edges are created, one passing inside and one outside the inpainting domain. The inner one tries to disappear but apparently gets stuck in a local energy well (local minimum) and does not seem to

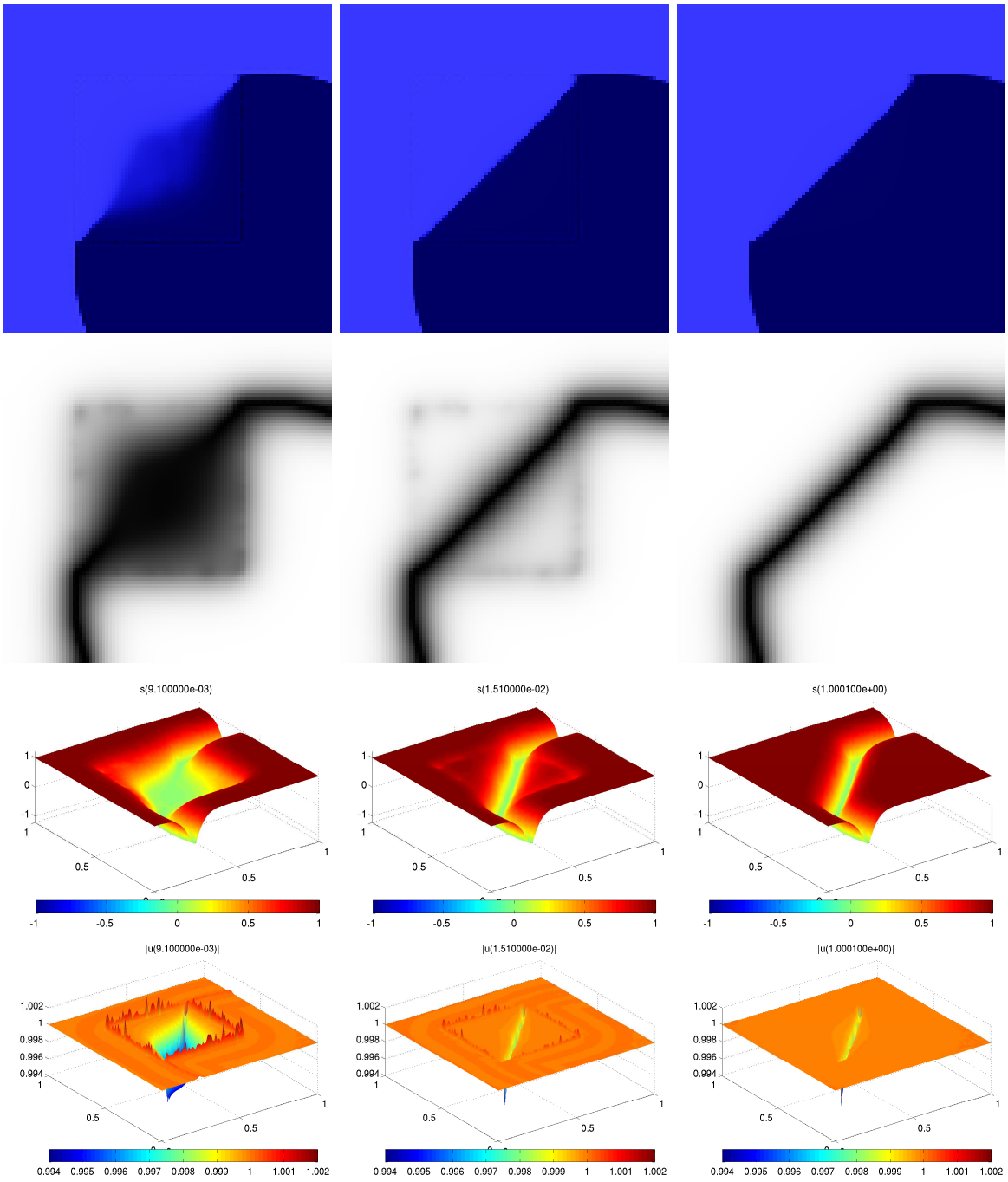

Figure 3.3: Example 3.1: $\beta = 0$; image (row 1), $|S|$ (row 2), S (row 3), and $|\mathbf{U}|$ (row 4) at times $t = 0.009, 0.015, 1$.

disappear even for times greater than 1. Indeed, the little circle we see in the lower right corner does not seem to be a computational artifact, but a stable radius r: The Euler term gives (for the full circle)

$$2\pi r \left(\alpha + \frac{\beta}{r^2} \right),$$

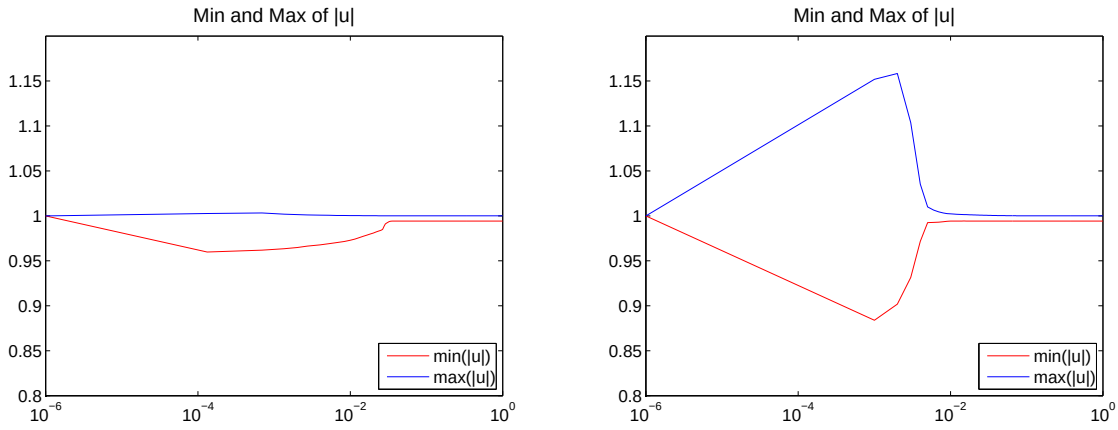

Figure 3.4: Example 3.1: min and max of $|\mathbf{U}|$ (x-logarithmic plots); $\beta = 0.2$ (left) and $\beta = 0$ (right).

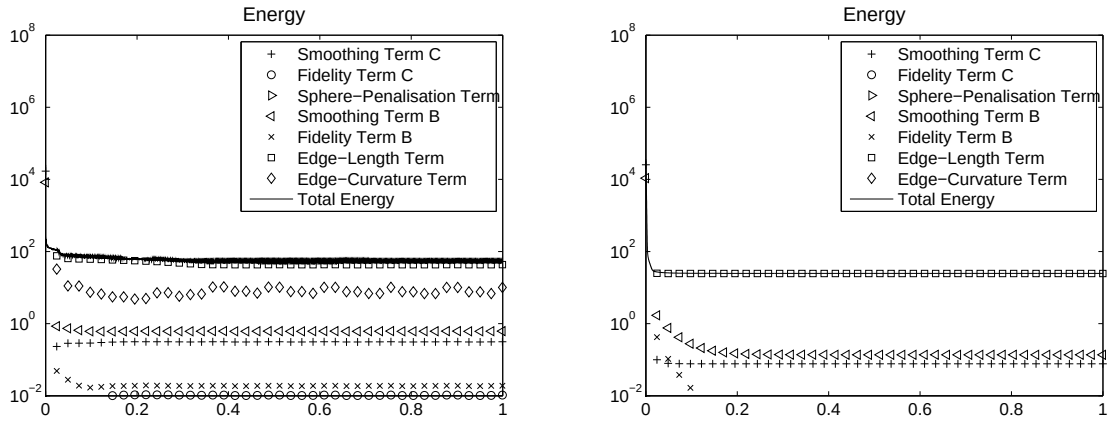

Figure 3.5: Example 3.1: Energy; $\beta = 0.2$ (left) and $\beta = 0$ (right), y-logarithmic plots.

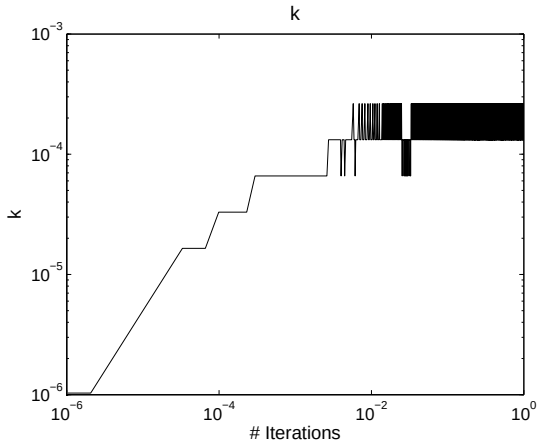

Figure 3.6: Example 3.1: $\beta = 0.2$; dynamic time-step (logarithmic plot).

whose derivative should be zero at minima, whence

$$r = \sqrt{\frac{\beta}{\alpha}} = 75^{-1/2} \approx 0.12,$$

which should be approximately the radius of that little circle. Changing the ratio between β and α by decreasing β or increasing α would make it smaller, but would also result in a less round reconstruction of the big circle, see Example 3.2.

Figure 3.4 shows the minimum and maximum of $|\mathbf{U}|$ over time for both cases. Interestingly, further decreasing δ_ε in the case of $\beta > 0$ moved $|\mathbf{U}|$ closer to 1, while there was almost no change in the case of $\beta = 0$. Figure 3.5 shows the energies for both cases ("Smoothing Term C" means the smoothing term of the chromaticity variable, while "Smoothing Term B" means the same for the brightness variable, etc.). The energy for $\beta = 0.2$ slightly oscillates, probably due to the oscillating dynamic time-step. Finally, Figure 3.6 shows the dynamic time-step, that results from our algorithm (for $\beta = 0.2$); in total, it took 5772 iterations to reach time 1. So, the average time step was approximately $2h^2$. But if we choose k so large from the beginning, the results turn out unpredictable and often chaotic.

3.6.3 Inpainting with Lagrange Multiplier, with and without Curvature

This example studies the same situation as Example 3.1, this time with Algorithm 3.5.1 (Lagrange multiplier).

Example 3.2. *Let $\Omega := (0,1)^2$ and $\mathbf{U}_0, V_0$ as in the right plot in Figure 3.1, and $S_0 \equiv 1$. The inpainting domain is filled with random data. Let $h = 1/100$, $\varepsilon = 10 * h$, and $k_\varepsilon = \varepsilon * 10^{-3}$. We compare the following sets of parameters:*

(1) *Let $\gamma = 50$, $\gamma_1 = 100$, $\alpha = 3$, $\beta = 0.02$, $\lambda = \lambda_1 = 5 * 10^5$, and $k_0 = h^3$ (the initial time step; the following steps are chosen dynamically according to Section 3.6.1), and $\mathrm{TOL} = 5 * 10^{-4}$.*

(2) *Let $\gamma = 50$, $\gamma_1 = 100$, $\alpha = 5$, $\beta = 0$, $\lambda = \lambda_1 = 10^6$, and $k_0 = h^3$, $\mathrm{TOL} = 10^{-3}$ (dynamic time-step as above).*

This example again shows that in some cases a curvature term is necessary to give natural inpainting results. As was to be expected, by lowering β much more than α, the reconstruction for $\beta > 0$ is not as round as in Example 3.1, but the spurious edge becomes much smaller. However, the dynamics take beyond time $t = 1$ to reach a stable state (for $\beta > 0$). The sphere-constraint is better preserved than with the penalisation algorithm, which qualitatively delivers the same dynamics when used with these same parameters.

Figure 3.7 shows iterates, edge sets (black means $S = 0$), phase functions S, and $|\mathbf{U}|$ at times $t = 0.17, 1, 2$ for inpainting with curvature, while Figure 3.8 shows the same at times $t = 0.0005, 0.08, 1$ for inpainting without curvature term (nothing changes beyond

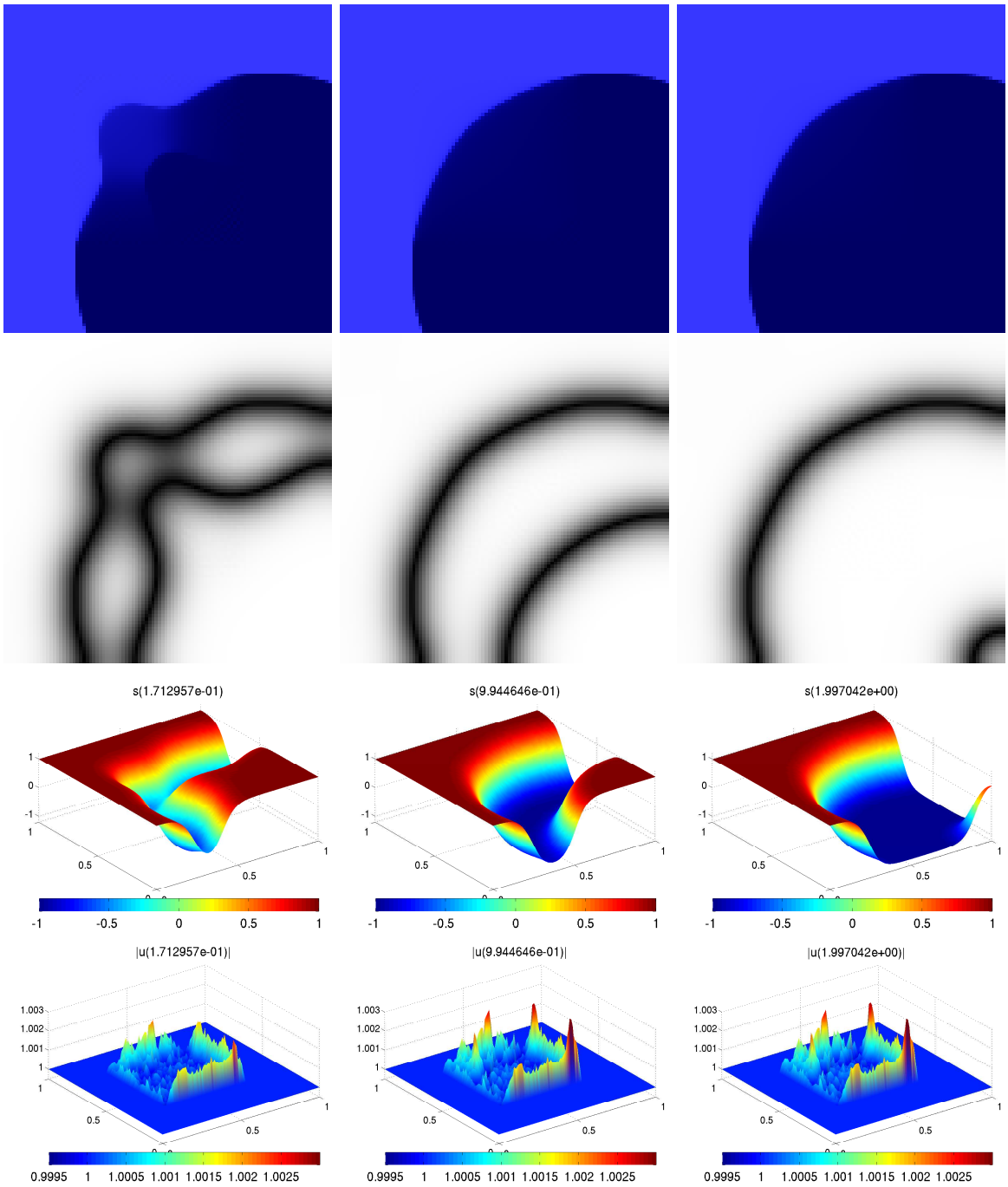

Figure 3.7: Example 3.2: $\beta = 0.02$; image (row 1), $|S|$ (row 2), S (row 3), and $|\mathbf{U}|$ (row 4) at times $t = 0.17, 1, 2$.

time 1). We see deviations from $|\mathbf{U}| = 1$ because of the fixed-point iteration and the discrete (i.e., approximate) Lagrange multiplier.

Figure 3.9 shows the minimum and maximum of $|\mathbf{U}|$ over time. We observe that the sphere-constraint is very well preserved, but errors seem to slightly "accumulate" over time. Figure 3.10 shows the energies, and Figure 3.11 shows the dynamic time-

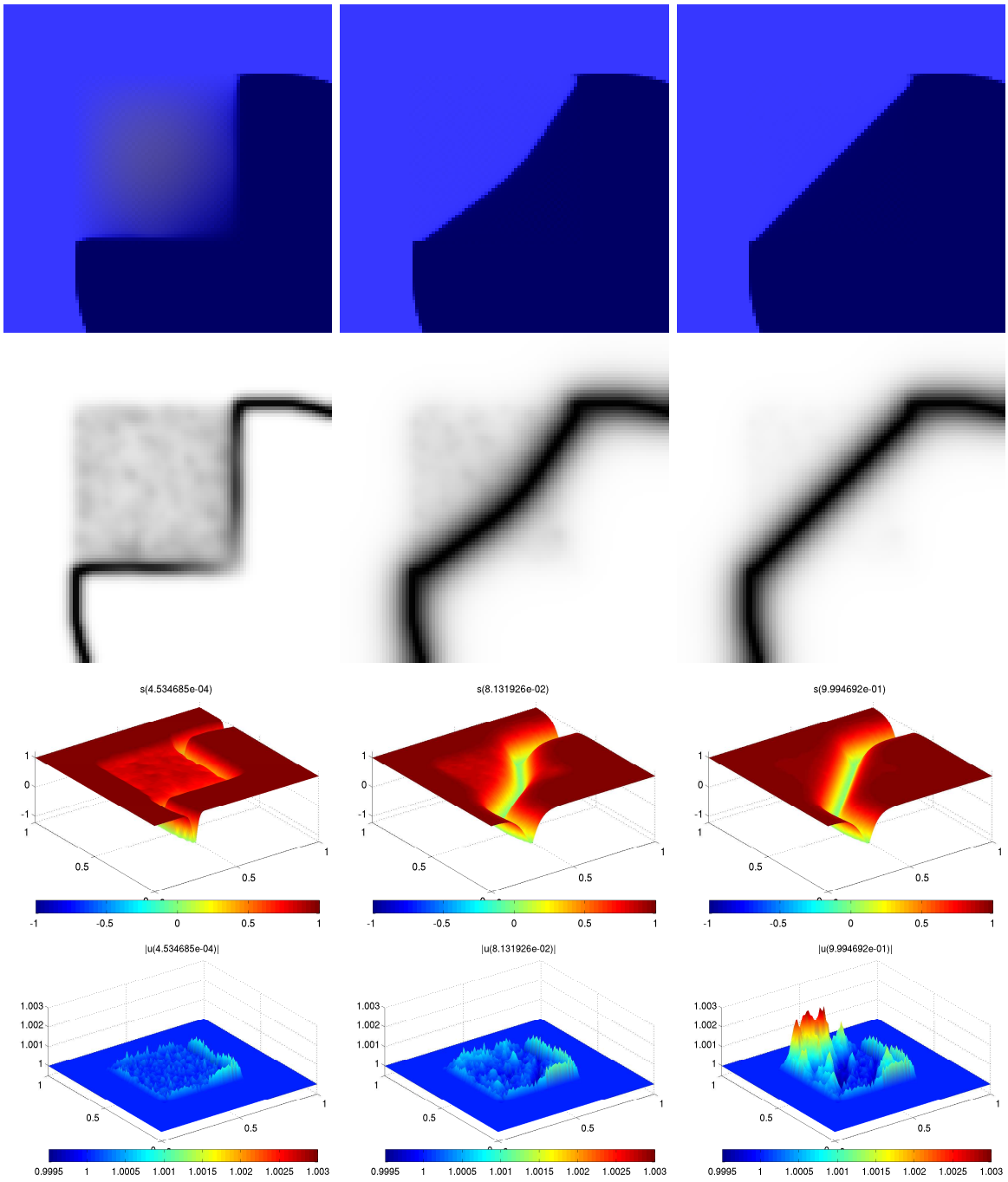

Figure 3.8: Example 3.2: $\beta = 0$; image (row 1), $|S|$ (row 2), S (row 3), and $|\mathbf{U}|$ (row 4) at times $t = 0.0005, 0.08, 1$.

step. It took only 1913 iterations to reach time $t = 2$, while for $\beta = 0$, it took 992 iterations to reach time $t = 1$ (even though the curvature term is zero when $\beta = 0$, we still observed better results with the dynamic time-step). The average time step was therefore approximately $10h^2$ (but the number of fixed-point iterations was 10 in most iterations).

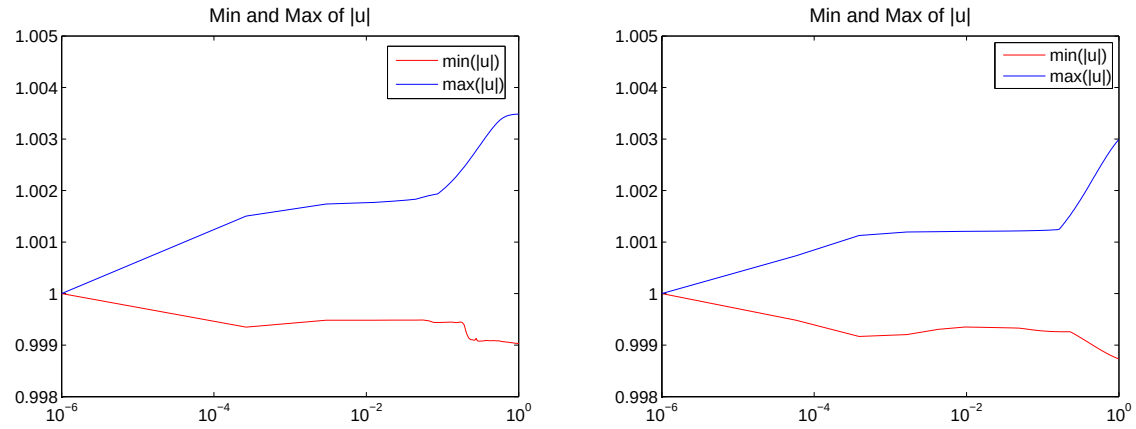

Figure 3.9: Example 3.2: min and max of $|\mathbf{U}|$; $\beta = 0.2$ (left) and $\beta = 0$ (right), x-logarithmic plots.

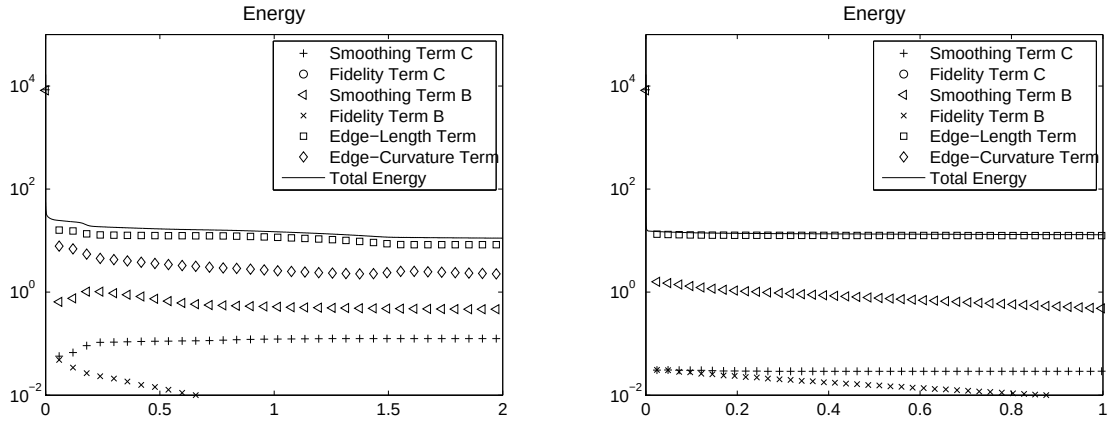

Figure 3.10: Example 3.2: Energy; $\beta = 0.02$ (left) and $\beta = 0$ (right), y-logarithmic plots.

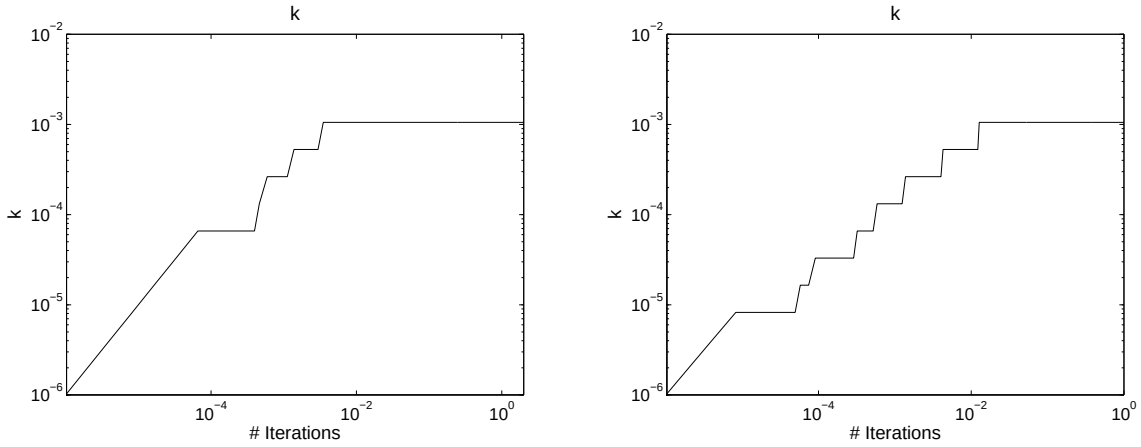

Figure 3.11: Example 3.2: Dynamic time-step; $\beta = 0.02$ (left) and $\beta = 0$ (right), logarithmic plots.

3.6.4 Inpainting with Penalisation, RGB vs CB

This example compares Algorithm 3.4.1 for CB and RGB for an initial image with CB noise.

Example 3.3. *Let* $\Omega := (0,1)^2$ *and use an initial image with CB noise, mainly in the chromaticity component:* $\mathbf{C} = \mathbf{C}_0 + 0.5 * \text{randn} * \mathbf{C}_0 \times [1,1,1]$, *and* $B = B_0 + 0.01 * \text{randn}$, *where* $\mathbf{C}$ *is projected to the sphere,* B *is cropped to lie in* $[0,1]$, *and* randn *are pseudo-random values drawn from the standard normal distribution. This is shown in Figure 3.12 (the even noisier square is again the inpainting domain). Choose* $S_0 \equiv 1$, $h = 1/100$, $\varepsilon = 10 * h$, *and* $k_\varepsilon = \varepsilon * 10^{-3}$. *We compare the following sets of parameters:*

(1) *Let* $\gamma = \gamma_1 = 90$, $\alpha = 15$, $\beta = .2$, $\lambda = 5 * 10^2$, $\lambda_1 = 5 * 10^5$, $\delta_\varepsilon = 10^{-4}$, *and* $k_0 = h^3$, *and* $\text{TOL} = 10^{-1}$.

(2) *Let* $\gamma = 100$, $\alpha = 20$, $\beta = .1$, $\lambda = 2e1$, *and* $k_0 = h^3$, *and* $\text{TOL} = 10^{-2}$.

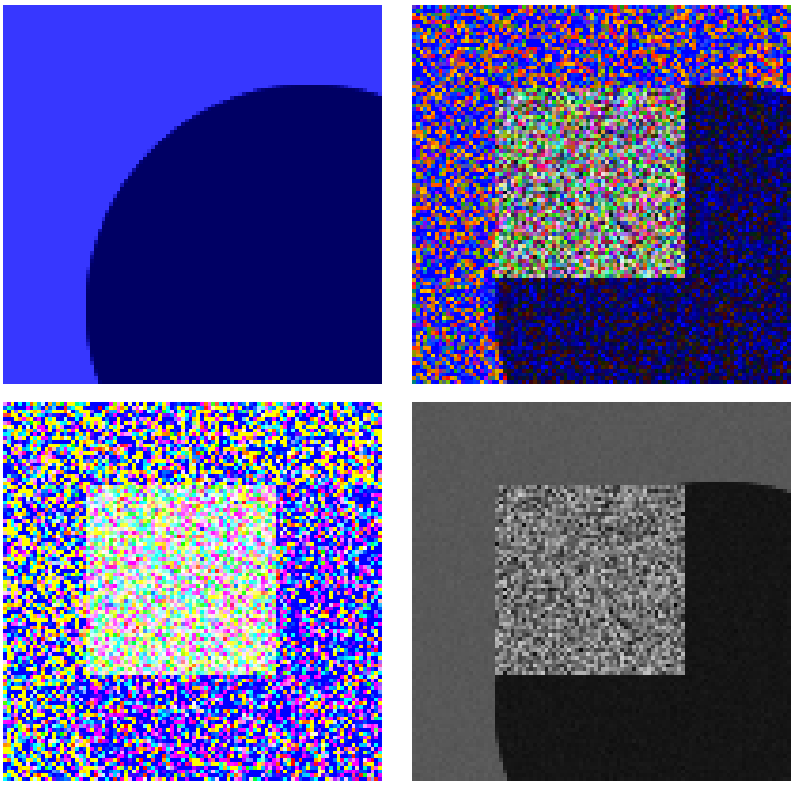

Figure 3.12: Example 3.3: Intact and degraded image (top left and right); degraded chromaticity and brightness (bottom left and right).

This example shows that with CB noise, RGB calculations cannot distinguish features from noise any more, while CB calculations can cope very well.

Figure 3.12 shows the intact image as well as the image we use as a starting point for inpainting.

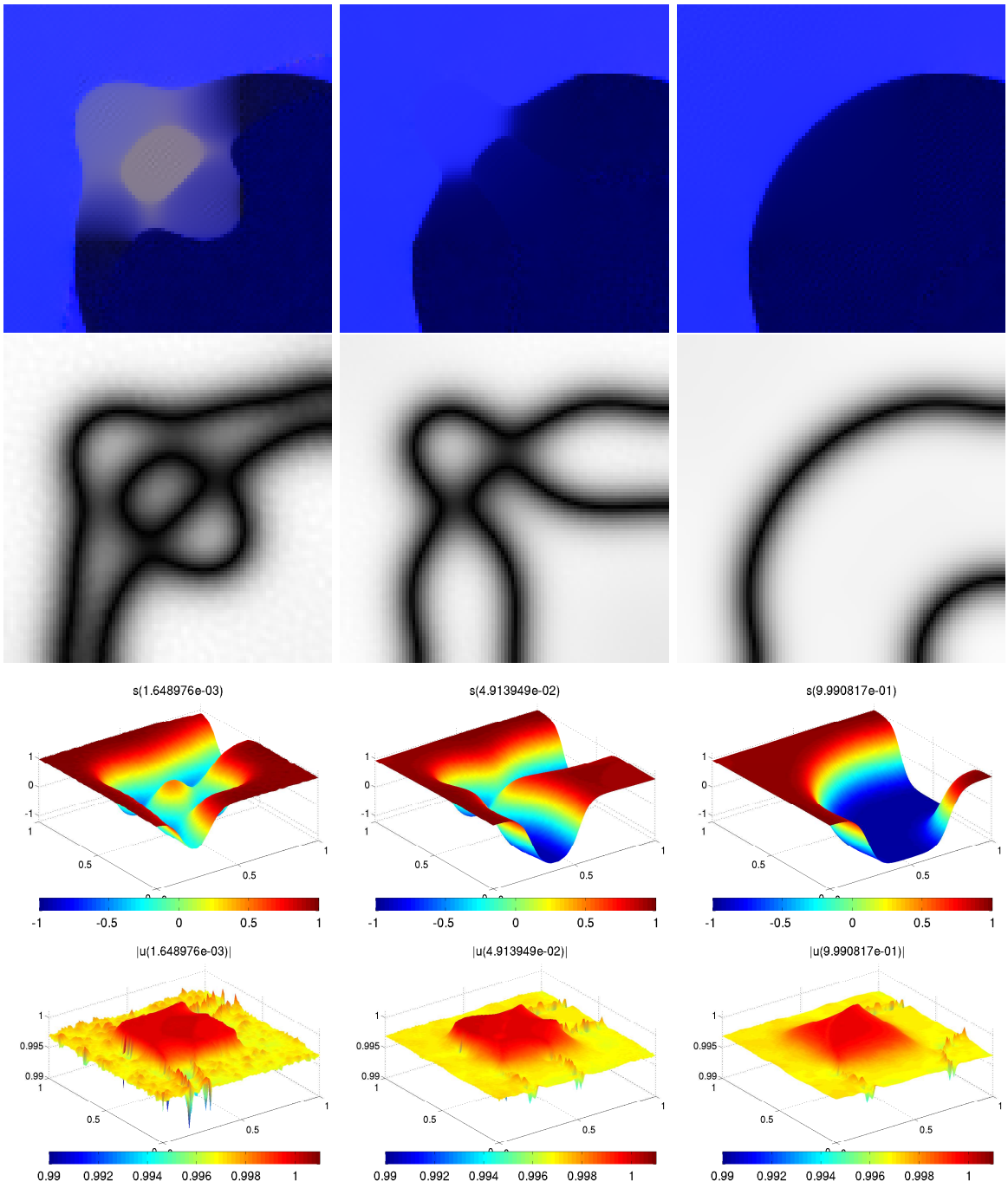

Figure 3.13: Example 3.3: CB; image (row 1), $|S|$ (row 2), S (row 3), and $|\mathbf{U}|$ (row 4) at times $t = 0.002, 0.05, 1$.

Figure 3.13 shows iterates, edge sets (black means $S = 0$), phase functions S, and $|\mathbf{U}|$ at times $t = 0.002, 0.05, 1$ for CB inpainting, and Figure 3.14 shows the minimum and maximum of $|\mathbf{U}|$ over time. The CB inpainting algorithm can reconstruct the image very well, noise notwithstanding, and the sphere constraint is still well preserved for the very small δ_ε we chose.

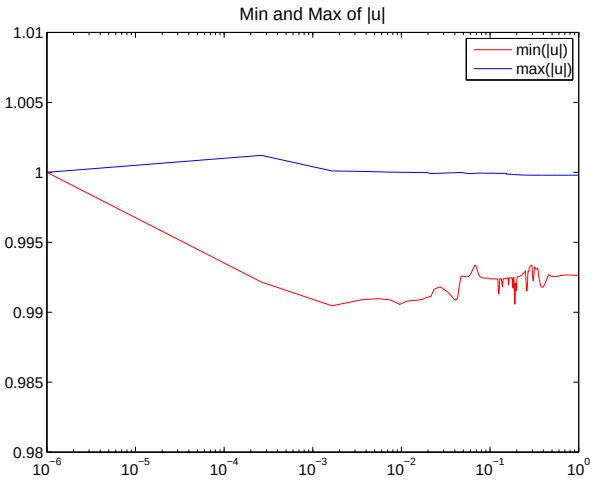

Figure 3.14: Example 3.3: min and max of $|\mathbf{U}|$ (CB, x-logarithmic plot).

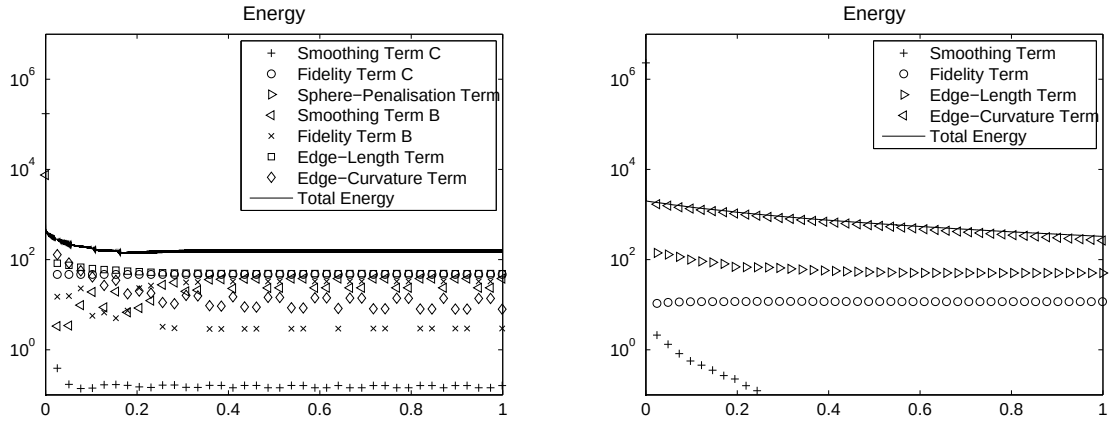

Figure 3.15: Example 3.3: Energy; CB (left) and RGB (right), y-logarithmic plots.

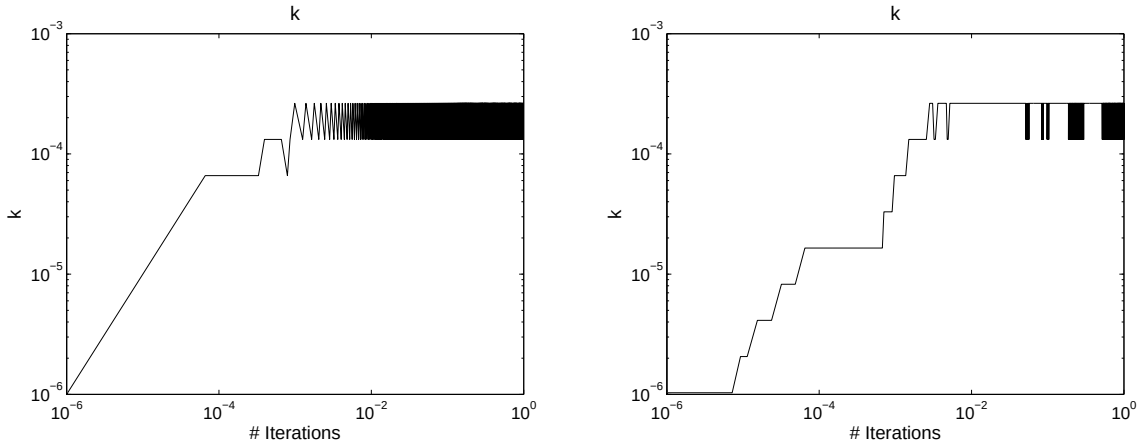

Figure 3.16: Example 3.3: dynamic time-step; CB (left) and RGB (right).

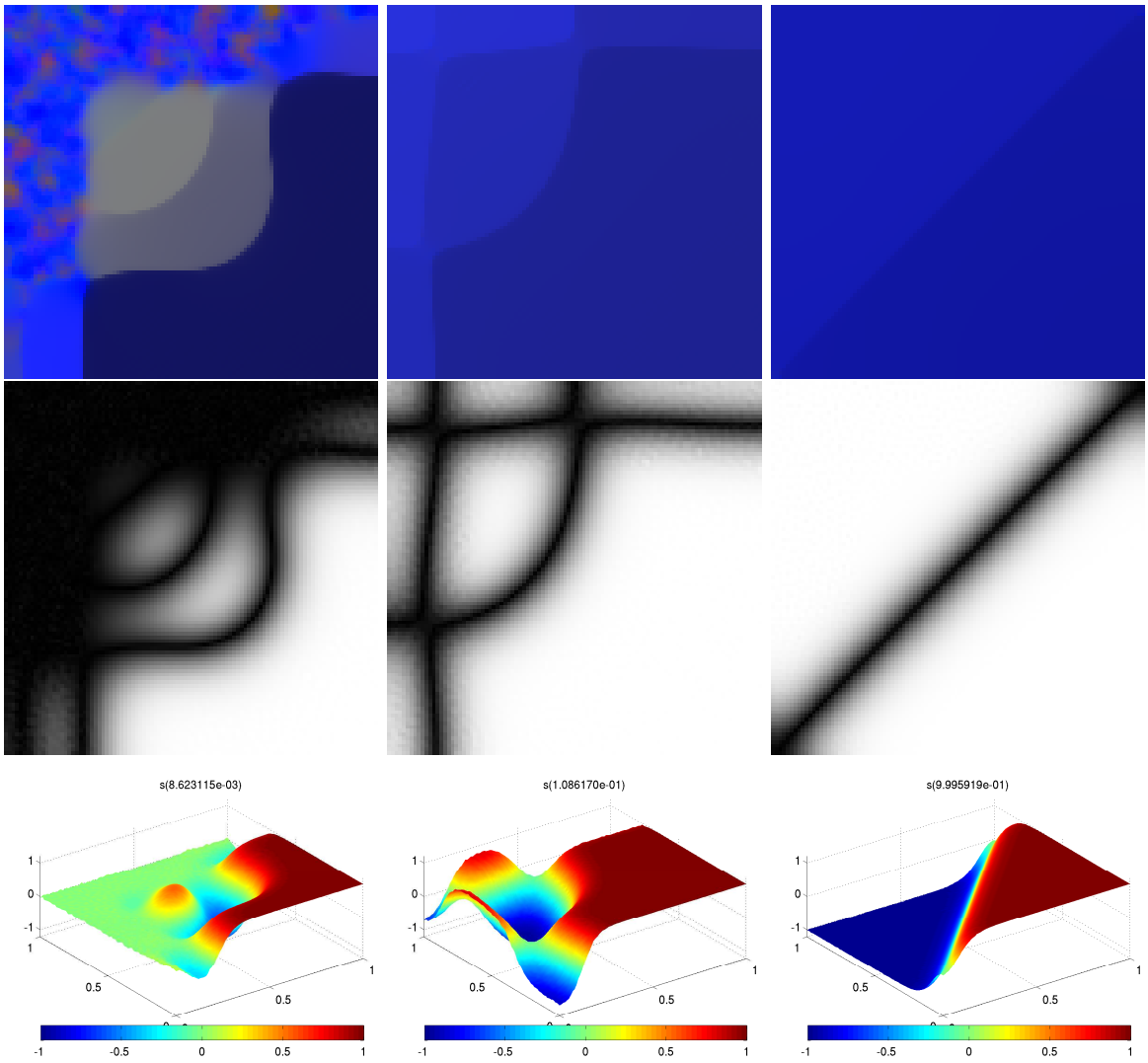

Figure 3.17: Example 3.3: RGB; image (row 1), $|S|$ (row 2), and S (row 3) at times $t = 0.009, 0.1, 1$.

Figure 3.15 shows the energies, with the energy for CB again showing slight oscillations due to the oscillating dynamic time step, which is shown in Figure 3.16. In total, it took 5064 iterations for the CB, and 4144 iterations for the RGB computations to reach time 1, so the average time step was again approximately $2h^2$.

Figure 3.17 shows iterates, edge sets, and phase functions S at times $t = 0.009, 0.1, 1$ for RGB inpainting. The reconstruction in this case has little to do with the original image. And taking a larger fidelity-parameter does not help either: In that case the noise just does not disappear.

3.6.5 Academic Example

In the following example, we only look at the vector-field $\mathbf{U}$. We are particularly interested in extreme situations like blow-up, so this section is not primarily about image

processing, and we work without a fidelity term (i.e., $\lambda = 0$), since it often blocks interesting dynamics. It will be interesting to compare the penalisation and Lagrange multiplier algorithm in this context.

Example 3.4. *Let* $\Omega := (0,1)^2$. *Choose* $\mathbf{U}_0$ *and* S_0 *according to Figure 3.18; i.e., arrows in the middle of* $\mathbf{U}_0$ *point down, while peripheral arrows point upwards, with smooth transitions between the two extremes, and* S_0 *equals the* z*-component of* $\mathbf{U}_0$. *Let* $\alpha = 10$, $\beta = 0.2$, $\gamma = 100$, $\varepsilon = 10 * h$, $k_\varepsilon = \varepsilon * 10^{-3}$, *and* $\mathrm{TOL} = 10^{-3}$.

We first use $h = 1/100$ *and compare results from the penalisation algorithm for* $\delta = 0.1$ *with those from the Lagrange multiplier algorithm (with the maximum number of allowable fixed-point iterations increased from* 10 *to* 20*).*

Then, we compare results from the penalisation algorithm for $h = 1/100$ *and* $\delta = 10^{-i}$, $i \in \{1, \ldots, 5\}$.

And finally, we compare results from the Lagrange multiplier algorithm for spatial discretisation $h \in \{0.01, 0.0125, 0.015, 0.02\}$.

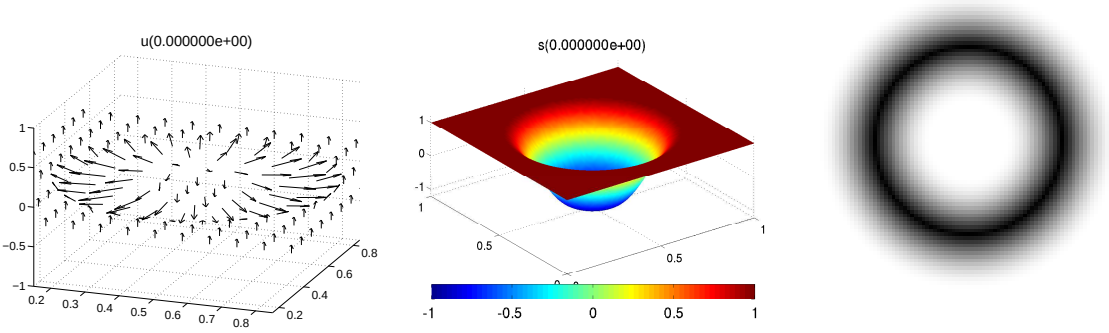

Figure 3.18: Example 3.4: $\mathbf{U}_0$, S_0 (which equals the z-component of $\mathbf{U}_0$), and $|S_0|$ (left to right).

This example shows that for highly nontrivial initial data, there can be considerable differences between the results returned from penalisation and Lagrange multipliers, depending on the choice of the penalisation parameter. Results for the algorithm using Lagrange multipliers appear more reliable.

Figure 3.18 shows the initial data, with a slight crop for $\mathbf{U}$ for better visibility.

The next five figures show the dynamics with the penalisation algorithm for $h = 1/100$ and $\delta = 0.1$: Figure 3.19 shows $\mathbf{U}$, $\mathbf{U}_z$, and $|\mathbf{U}|$ at times $t = 0.001, 0.015, 1$, and Figure 3.20 shows S and $|S|$ at the same points in time. Iterates $\mathbf{U}$ initially show steepening of gradients, but then move to a constant vector field fairly quickly, even though it might be expected that $S = 0$ at places of high gradients of $\mathbf{U}$ should allow the central vectors of $\mathbf{U}$ to keep pointing in the opposite direction of its peripheral peers. The reason for this is revealed in the bottom row of Figure 3.19: Because the sphere-constraint is only slightly penalised, "tunnelling" happens; i.e., the vectors, particularly in the centre, get shorter until they can flip directions without too much cost, see also Figure 3.21, which shows the minimum and maximum of $|\mathbf{U}|$ over time. What remains

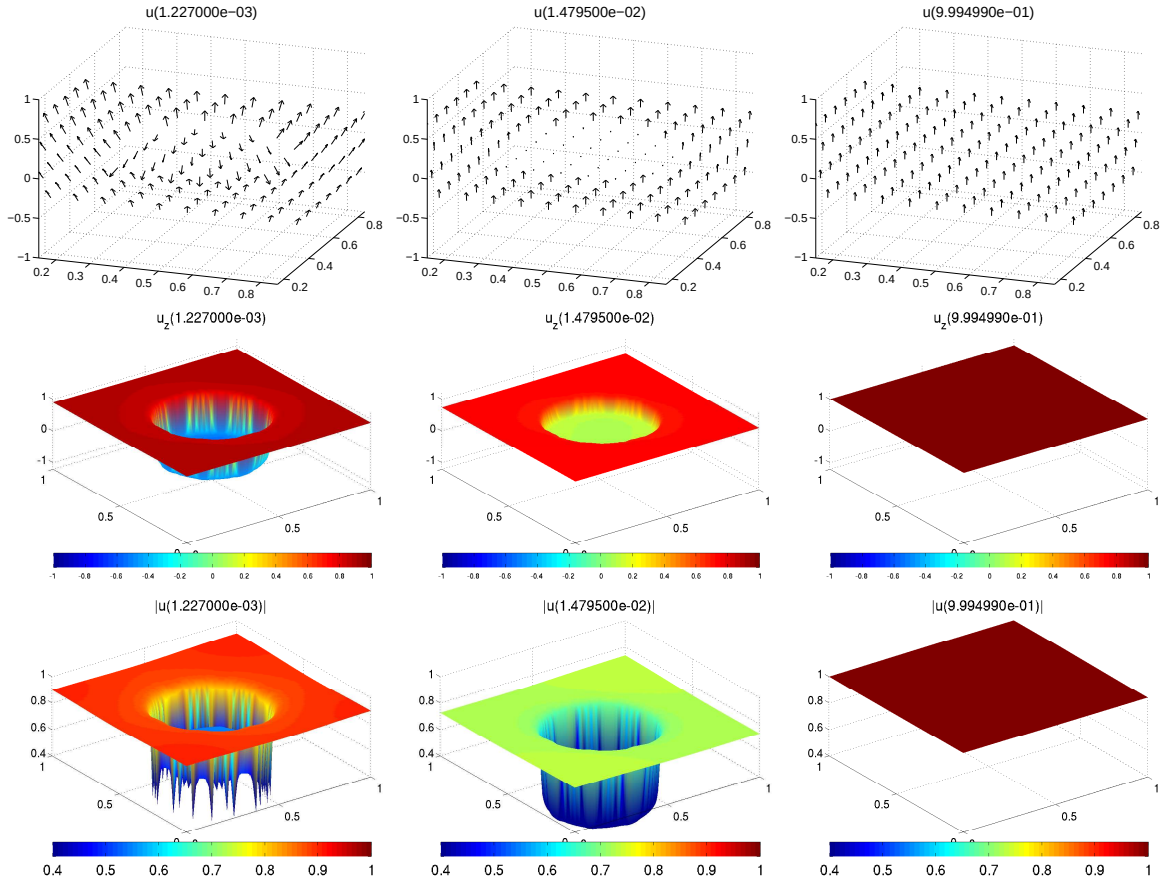

Figure 3.19: Example 3.4: Penalisation, $\delta = 0.1$; $\mathbf{U}$, $\mathbf{U}_z$, and $|\mathbf{U}|$ at times $t = 0.001, 0.015, 1$.

at the end, is a set of interfaces $S = 0$ with a stable radius (which happens when the length and curvature term reach a balance — without a curvature term, there would be no such thing, and all interfaces would vanish). Figure 3.22 shows $W^{1,\infty}(\mathbf{U})$, in a linear plot (left) and an x-logarithmic one, in order to show the effect of the initial steepening of gradients. And finally, Figure 3.23 shows the energy and the dynamic time-step k (the total number of steps is 6253).

The following seven figures show the dynamics with the Lagrange multiplier algorithm for $h = 1/100$ with many images, due to the complexity of the dynamics. For this example, we increase the maximum number of allowable fixed-point iterations from 10 to 20, due to the complexity of the data. This results in the sphere-constraint being extremely well preserved; indeed the global in time and space minimum and maximum of $|\mathbf{U}|$ are 0.9997 and 1.0017, respectively, see Figure 3.28 (left). We therefore omit the spatial plots of $|\mathbf{U}|$. Note that increasing the number of fixed-point iterations for the penalised algorithm would not change much, since the sphere-constraint is the critical point here, observance of which does not improve with the number of fixed-point iterations in the penalised case.

Figure 3.24 shows $\mathbf{U}$, and Figure 3.25 shows the z-component of $\mathbf{U}$, both at times $t = 0.003, 0.05, 0.26, 0.62, 0.63, 1$, while Figures 3.26 and 3.27 show S and $|S|$ at the same points in time. In image one in Figures 3.24 and 3.25, we again see a steepening of

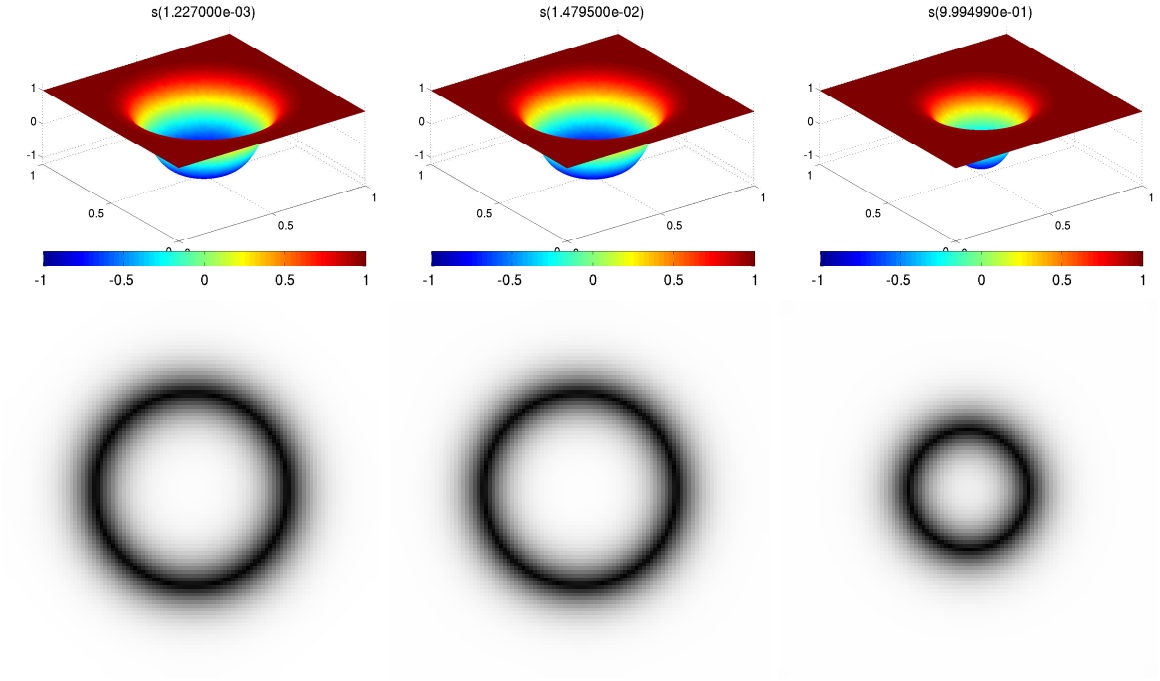

Figure 3.20: Example 3.4: Penalisation, $\delta = 0.1$; S and $|S|$ at times $t = 0.001, 0.015, 1$.

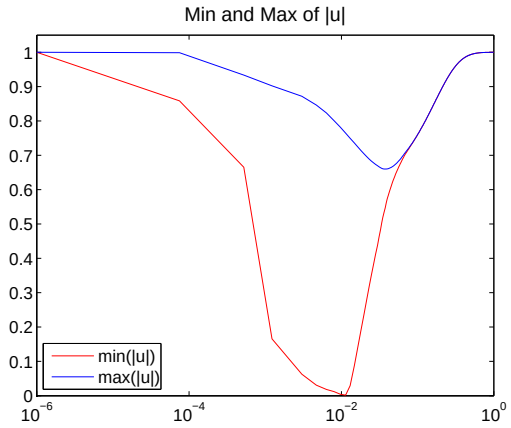

Figure 3.21: Example 3.4: Penalisation, $\delta = 0.1$; min and max of $|\mathbf{U}|$ over time.

gradients compared to the initial data of Figure 3.18, which corresponds to an increase in $W^{1,\infty}(\mathbf{U})$ (Figure 3.28, right). In image two we see that a first ring of arrows flips half-way, corresponding to a double-ring in S (image one and two in Figures 3.26 and 3.27) and a sharp decline in $W^{1,\infty}(\mathbf{U})$. Then, this first row of arrows flips completely, resulting in maximally sharp interfaces (image three in Figures 3.24 and 3.25) and maximal $W^{1,\infty}(\mathbf{U}) = 2/h$, which is kept until $t \approx 0.6$, when suddenly all remaining arrows in the centre flip (this is probably a result of numerical imprecisions), accompanied by a sudden decrease in $W^{1,\infty}(\mathbf{U})$ and in the energy (Figure 3.29, left). Now, the interfaces S are no longer constrained by jumps in $\nabla\mathbf{U}$ and therefore move to a stable radius. Finally,

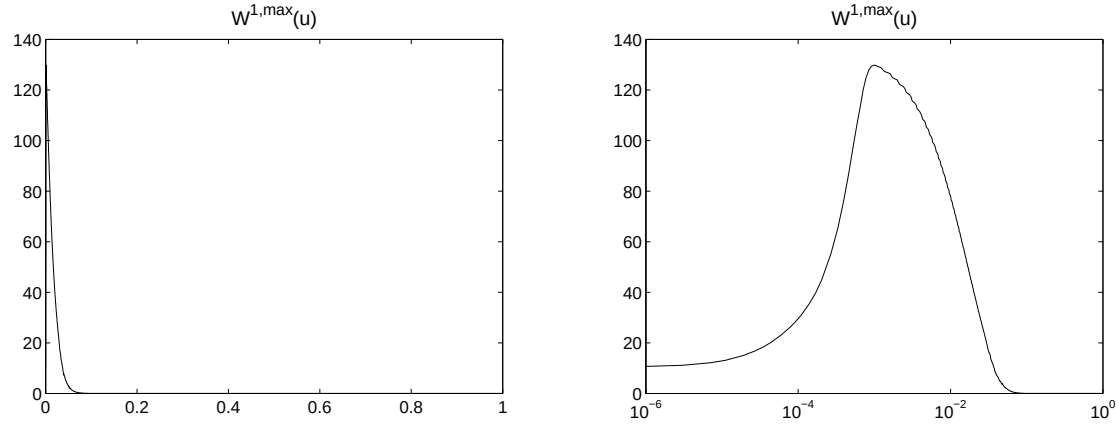

Figure 3.22: Example 3.4: Penalisation, $\delta = 0.1$; $W^{1,\infty}(\mathbf{U})$, linear and x-logarithmic plot.

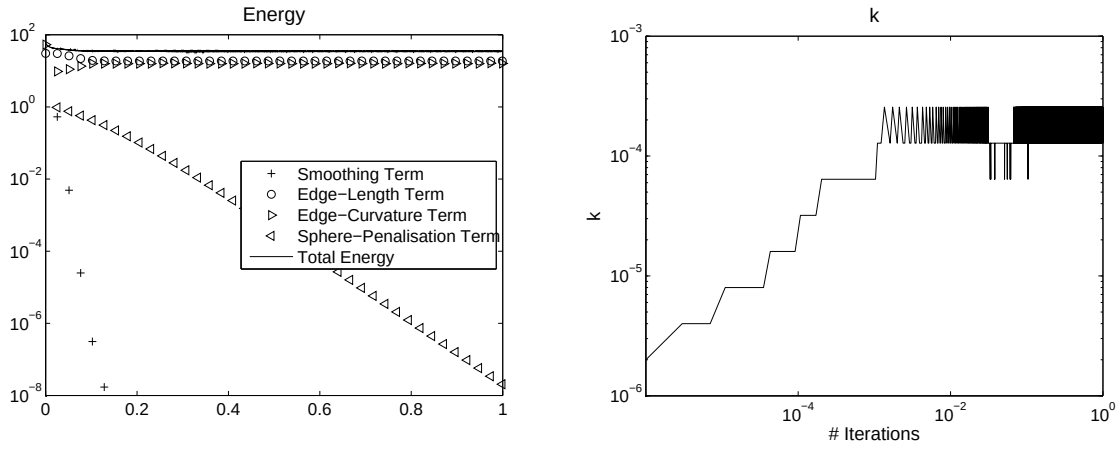

Figure 3.23: Example 3.4: Penalisation, $\delta = 0.1$; energy and dynamic time-step k (y-logarithmic and fully logarithmic plot, respectively).

the right image in Figure 3.29 again shows the dynamic time-step k (the total number of steps is 7596).

For both the penalisation and the Lagrange multiplier algorithm, $W^{1,\infty}(S)$ (not shown here) behaves similarly: It starts with a value of about 10 and increases to about 14 (which it reaches around $t = 0.15$ for the Lagrange multiplier algorithm, and somewhat sooner for the penalisation). After that it does not change significantly any more. Also, $W^{1,\infty}(S)$ changes only slightly with h in the studies below.

Then, Figures 3.30 and 3.31 compare results from the penalisation algorithm with $\delta = 10^{-i}$ for $i \in \{1, \ldots, 5\}$. The left plot of Figure 3.30 shows $\min\{|\mathbf{U}|\}$, displaying strong signs of tunnelling for $\delta = 0.1, 0.01$ and even 0.001, and oscillations for $\delta = 10^{-5}$; only in the case of $\delta = 10^{-4}$ is $|\mathbf{U}|$ reasonably well preserved. The right plot shows the dynamic time-step k, which becomes very small for $\delta = 10^{-5}$. Figure 3.31 shows $W^{1,\infty}(\mathbf{U})$ and the energy. Both seem far off for $\delta = 0.1$ and $\delta = 0.01$. In the case of $\delta = 10^{-3}$ and $\delta = 10^{-4}$, the plots look reasonable, with $\delta = 10^{-4}$ looking very similar to what the Lagrange multiplier algorithm delivered (c.f. Figures 3.28 and 3.29), while $\delta = 10^{-5}$ leads

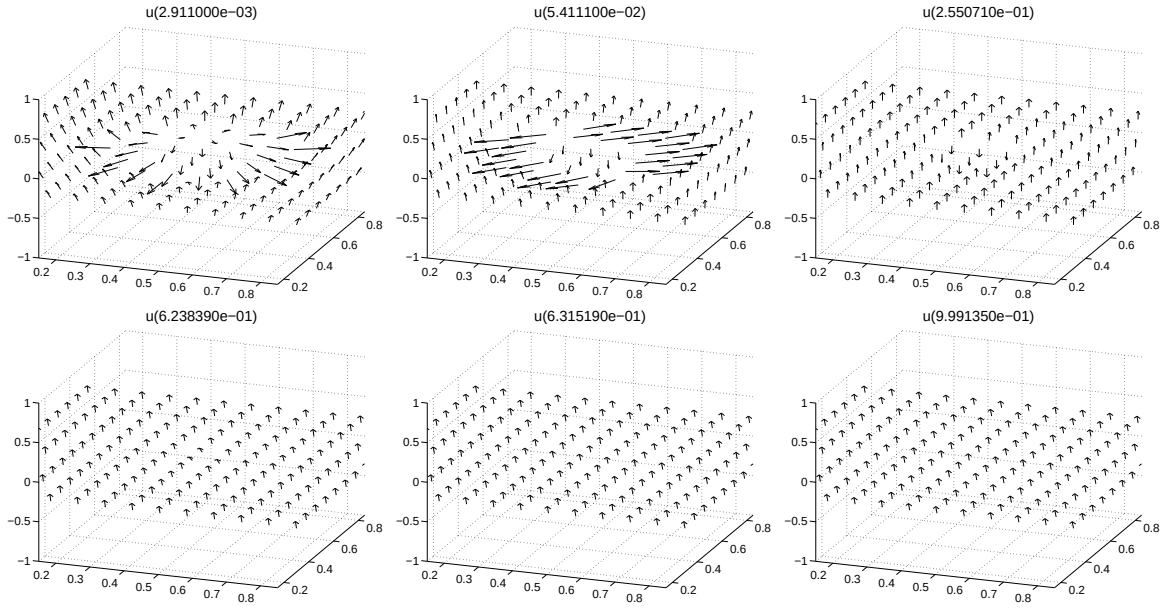

Figure 3.24: Example 3.4: Lagrange multiplier; $\mathbf{U}$ at times $t = 0.003, 0.05, 0.26, 0.62, 0.63, 1$.

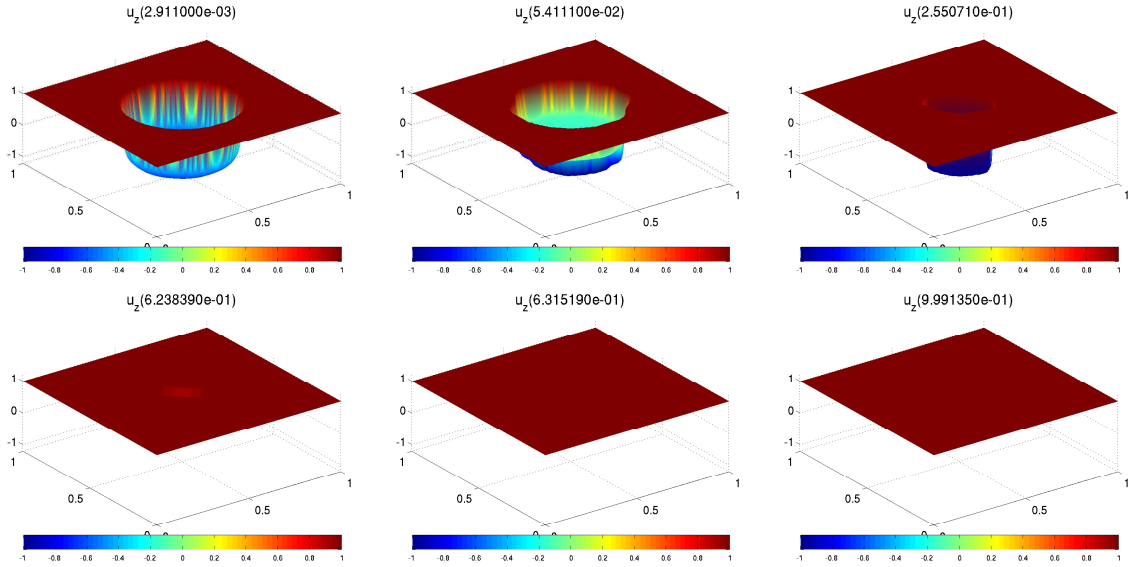

Figure 3.25: Example 3.4: Lagrange multiplier; $\mathbf{U}_z$ at times $t = 0.003, 0.05, 0.26, 0.62, 0.63, 1$.

to very strong oscillations. Indeed, the dynamics of $\mathbf{U}$ and S are close to those from the Lagrange multiplier algorithm only for $\delta = 10^{-4}$ (not shown here).

The last two figures, namely Figures 3.32 and 3.33, show $W^{1,\infty}(\mathbf{U})$, energy and dynamic time-step for the Lagrange multiplier algorithm for various spatial discretisations $h \in \{0.01, 0.0125, 0.015, 0.02\}$. The dynamics are qualitatively similar for different h, but the finer h, the longer it takes for the "flipping" to take place. Figure 3.32 confirms that,

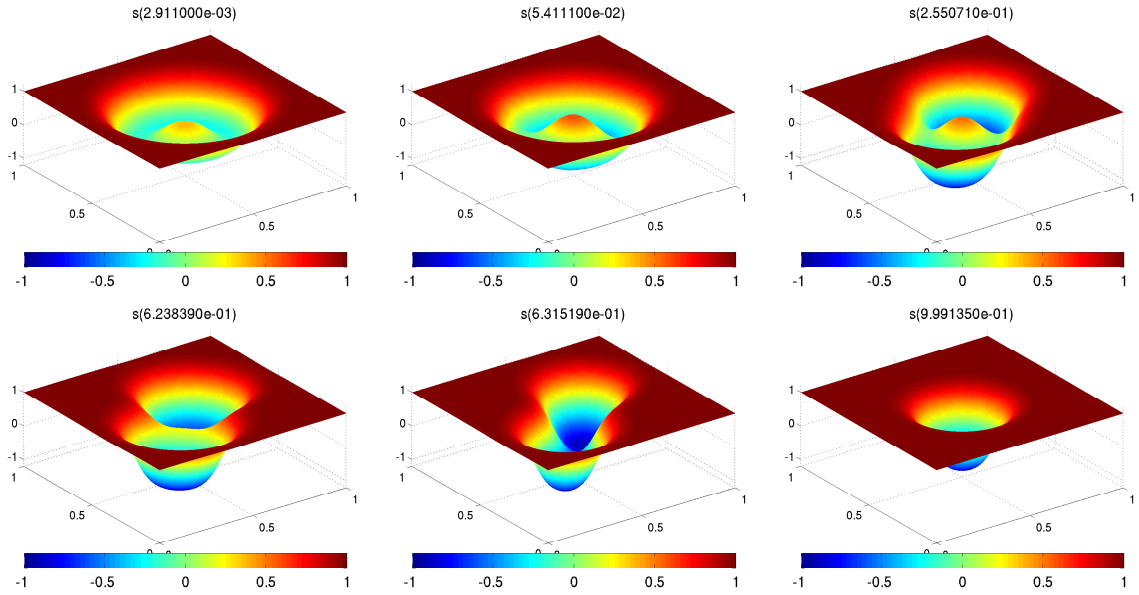

Figure 3.26: Example 3.4: Lagrange multiplier; S at times $t = 0.003, 0.05, 0.26, 0.62, 0.63, 1.$

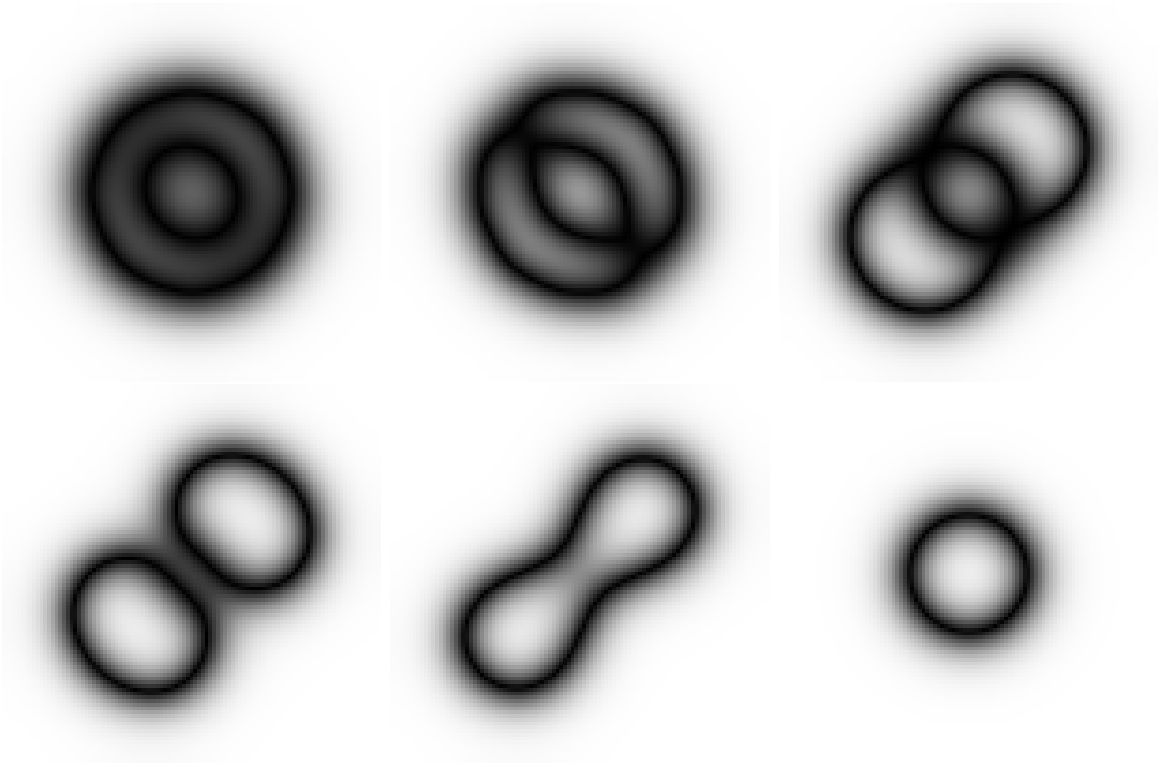

Figure 3.27: Example 3.4: Lagrange multiplier; $|S|$ at times $t = 0.003, 0.05, 0.26, 0.62, 0.63, 1.$

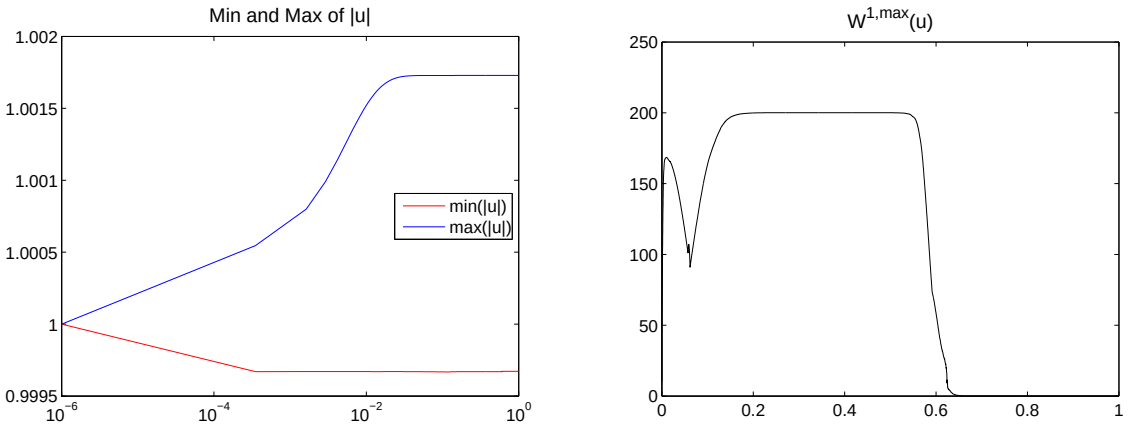

Figure 3.28: Example 3.4: Lagrange multiplier; min and max of $|\mathbf{U}|$ (x-logarithmic) and $W^{1,\infty}(\mathbf{U})$ over time.

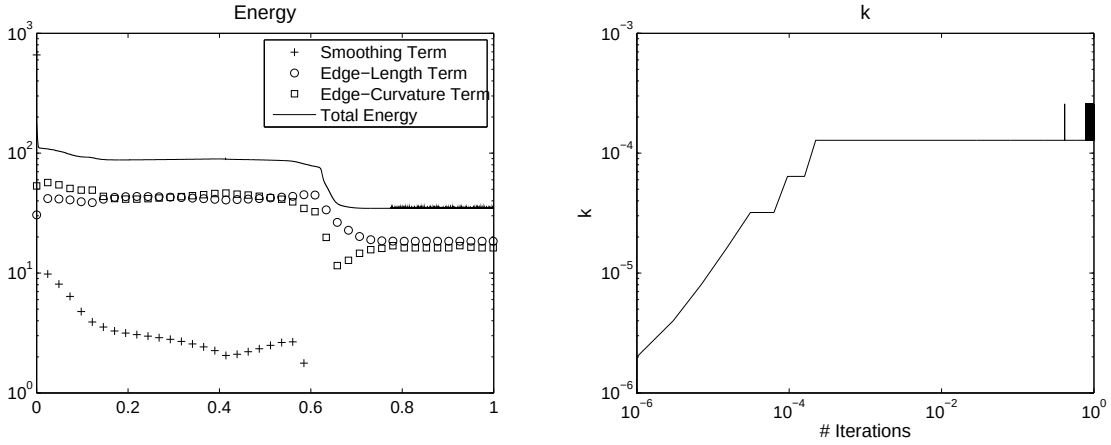

Figure 3.29: Example 3.4: Lagrange multiplier; energy and dynamic time-step k (y-logarithmic and fully logarithmic plot, respectively).

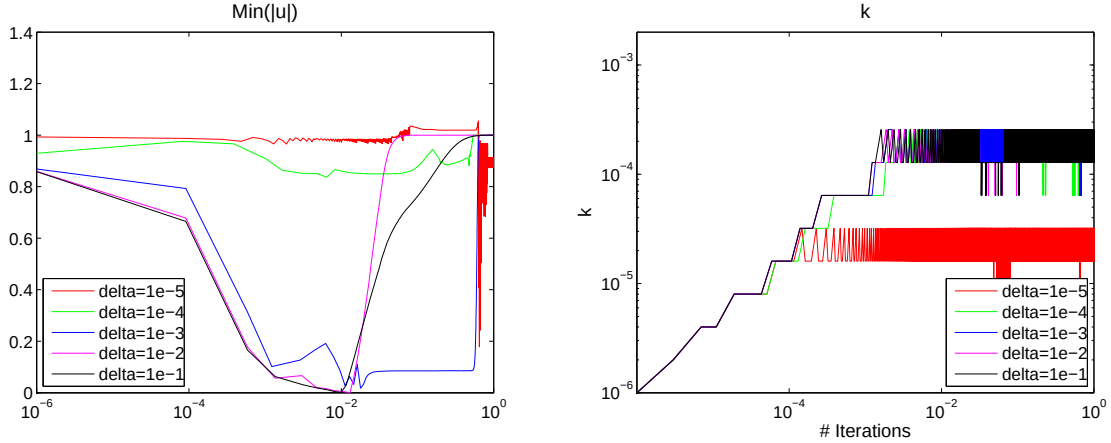

Figure 3.30: Example 3.4: Penalisation; $\min\{|\mathbf{U}|\}$ and dynamic time-step k for $\delta = 10^{-i}$, $i \in \{1, \ldots, 5\}$ (x-logarithmic and fully logarithmic, respectively).

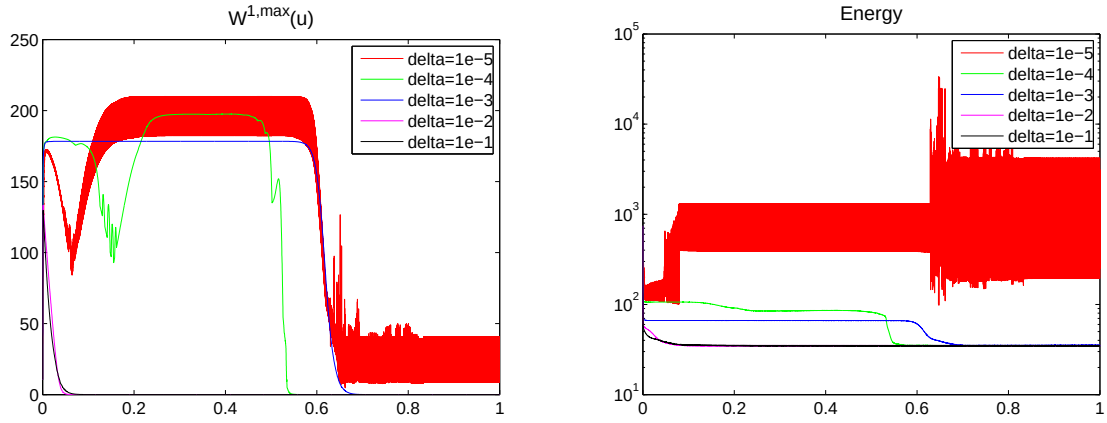

Figure 3.31: Example 3.4: Penalisation; $W^{1,\infty}(\mathbf{U})$ and energy for $\delta = 10^{-i}$, $i \in \{1, \ldots, 5\}$ (y-logarithmic for the energy).

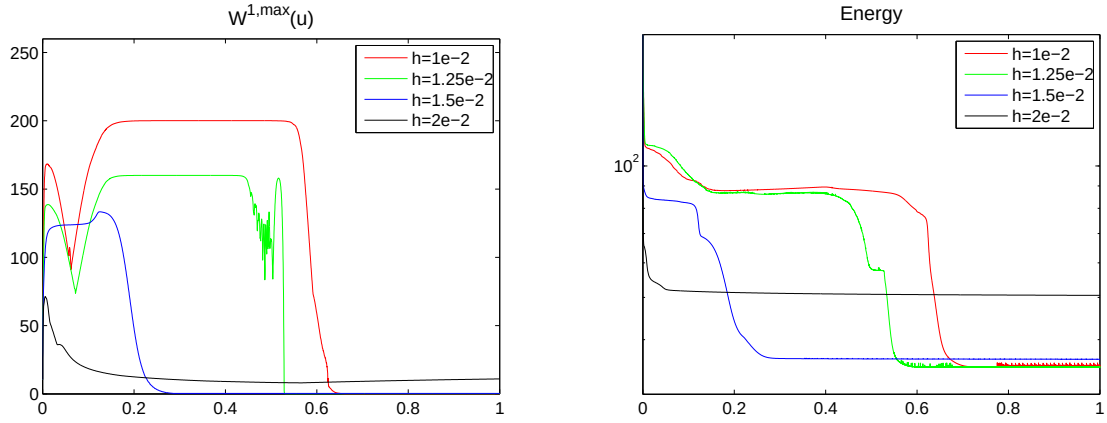

Figure 3.32: Example 3.4: Lagrange multiplier; $W^{1,\infty}(\mathbf{U})$ and energy for $h \in \{0.01, 0.0125, 0.015, 0.02\}$ (y-logarithmic for the energy).

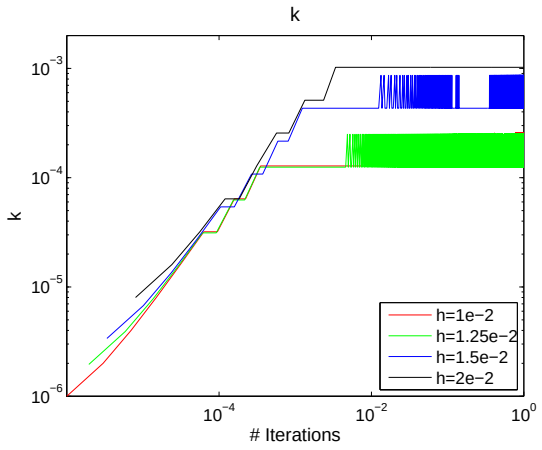

Figure 3.33: Example 3.4: Lagrange multiplier; Dynamic time-step k for $h \in \{0.01, 0.0125, 0.015, 0.02\}$, logarithmic plot.

indeed, $W^{1,\infty}(\mathbf{U})$ in the case of our Lagrange multiplier algorithm is bounded by $2/h$, and thus motivates blow-up behaviour ($W^{1,\infty}(\mathbf{U}_0) \approx 10$ for all h).

3.7 Conclusion

In this chapter, we proposed two strategies for the L^2 Mumford-Shah-Euler flow for unit vector fields $\mathbf{u} : \Omega \subset \mathbb{R}^d \to \mathbb{S}^{m-1}$, based on first-order finite elements. Both strategies use elliptic approximations to the original functional, based on the Ambrosio-Tortorelli functional with modified length and added curvature term, corresponding to the De Giorgi approximation of the Willmore energy.

The first strategy uses a penalisation approach, adding a Ginzburg-Landau term to the energy; so iterates $\mathbf{U}$ only approximate the sphere constraint. For the resulting Algorithm 3.4.1, we can show existence of solutions, an energy principle, and strong convergence of iterates $\nabla\mathbf{U}$ in L^2, as well as weak convergence of iterates $\widetilde{\Delta}_h S$ in L^2, both in space and time. This allows us to show that iterates indeed converge to a weak solution $(\mathbf{u}, s)$ as in Definition 3.3.1. We only need a weak coupling $k = o(h)$ for this convergence, but a stronger coupling $k \leq C\varepsilon^{-1}h^4$ for existence of discrete solutions.

The second strategy uses a discrete Lagrange multiplier that enforces the sphere constraint exactly in the continuous case. The necessity to approximate the Lagrange multiplier in the discrete setting, and to solve the resulting nonlinear equation through a fixed-point iteration, again slightly distorts $|\mathbf{U}|$. For the resulting Algorithm 3.5.1, we need a strong coupling $k \leq \widetilde{C} \max\left\{h^{d+2}, \varepsilon^{-1}h^4\right\}$ to show existence of solutions, an energy principle, and weak L^2 convergence of iterates $\nabla\mathbf{U}$ and $\widetilde{\Delta}_h S$. Identifying the limit $(\mathbf{u}, s)$ as a weak solution as in Definition 3.3.1 is still an open problem.

We observe that the sphere-consistent algorithm gives results that appear somewhat more reliable and better preserve the sphere constraint, especially for complicated initial data.

Acknowledgements

For this chapter, we thank Q. Du (Penn State) and M. Rumpf (U Bonn) for helpful discussions.

Bibliography

[1] F. Alouges. A new algorithm for computing liquid crystal stable configurations: the harmonic mapping case. *SIAM J. Numer. Anal.*, 34(5):1708–1726, 1997.

[2] F. Alouges, S. Conti, A. DeSimone, and Y. Pokern. Energetics and switching of quasi-uniform states in small ferromagnetic particles. *M2AN Math. Model. Numer. Anal.*, 38(2):235–248, 2004.

[3] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[4] L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Comm. Pure Appl. Math.*, 43(8):999–1036, 1990.

[5] L. Ambrosio and V. M. Tortorelli. On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital. B (7)*, 6(1):105–123, 1992.

[6] L. Baňas, A. Prohl, and R. Schätzle. Finite element approximations of harmonic map heat flows and wave map into spheres of nonconstant radii. Preprint, URL: http://na.uni-tuebingen.de/preprints.shtml, June 2008.

[7] J. W. Barrett, H. Garcke, and R. Nürnberg. A parametric finite element method for fourth order geometric evolution equations. *J. Comput. Phys.*, 222(1):441–462, 2007.

[8] S. Bartels. A posteriori error analysis for time-dependent Ginzburg-Landau type equations. *Numer. Math.*, 99(4):557–583, 2005.

[9] S. Bartels. Robust a priori error analysis for the approximation of degree-one Ginzburg-Landau vortices. *M2AN Math. Model. Numer. Anal.*, 39(5):863–882, 2005.

[10] S. Bartels. Stability and convergence of finite-element approximation schemes for harmonic maps. *SIAM J. Numer. Anal.*, 43(1):220–238 (electronic), 2005.

[11] S. Bartels, C. Lubich, and A. Prohl. Convergent discretization of heat and wave map flows to spheres using approximate discrete Lagrange multipliers. *Math. Comp.*, 78(267):1269–1292, 2009.

[12] S. Bartels and A. Prohl. Constraint preserving implicit finite element discretization of harmonic map flow into spheres. *Math. Comp.*, 76(260):1847–1859 (electronic), 2007.

[13] S. Bartels and A. Prohl. Stable discretization of scalar and constrained vectorial Perona-Malik equation. *Interfaces Free Bound.*, 9(4):431–453, 2007.

[14] G. Bellettini and A. Coscia. Discrete approximation of a free discontinuity problem. *Numer. Funct. Anal. Optim.*, 15(3-4):201–224, 1994.

[15] G. Bellettini and R. March. An image segmentation variational model with free discontinuities and contour curvature. *Math. Models Methods Appl. Sci.*, 14(1):1–45, 2004.

[16] G. Bellettini and M. Paolini. Approssimazione variazionale di funzionali con curvatura. *Seminario di Analisi Matematica, Dipartimento di Matematica dell'Università di Bologna*, pages 87–97, 1993.

[17] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[18] F. Bethuel. On the singular set of stationary harmonic maps. *Manuscripta Math.*, 78(4):417–443, 1993.

[19] G. Birkhoff and C. R. De Boor. Piecewise polynomial interpolation and approximation. In *Approximation of Functions (Proc. Sympos. General Motors Res. Lab., 1964)*, pages 164–190. Elsevier Publ. Co., Amsterdam, 1965.

[20] B. Bourdin. Image segmentation with a finite element method. *M2AN Math. Model. Numer. Anal.*, 33(2):229–244, 1999.

[21] A. Braides. *Approximation of free-discontinuity problems*, volume 1694 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1998.

[22] A. Braides. Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[23] A. Braides and G. Dal Maso. Non-local approximation of the Mumford-Shah functional. *Calc. Var. Partial Differential Equations*, 5(4):293–322, 1997.

[24] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2002.

[25] S. Burke, C. Ortner, and E. Süli. An adaptive finite element approximation of a variational model of brittle fracture. *OxMOS Preprint No. 16*, 2008.

[26] M. C. Calderer, D. Golovaty, F.-H. Lin, and C. Liu. Time evolution of nematic liquid crystals with variable degree of orientation. *SIAM J. Math. Anal.*, 33(5):1033–1047 (electronic), 2002.

[27] M. Carriero and A. Leaci. S^k-valued maps minimizing the L^p-norm of the gradient with free discontinuities. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 18(3):321–352, 1991.

[28] T. Cecil, S. Osher, and L. Vese. Numerical methods for minimization problems constrained to S^1 and S^2. *J. Comput. Phys.*, 198(2):567–579, 2004.

[29] A. Chambolle. Inverse problems in image processing and image segmentation: some mathematical and numerical aspects. In C. E. Chidume, editor, *Mathematical Problems in Image Processing*. ICTP, December 2000. ICTP Lecture Notes Series, Vol 2, URL: http://publications.ictp.it/lns/vol2.html.

[30] T. Chan and J. Shen. Variational restoration of nonflat image features: models and algorithms. *SIAM J. Appl. Math.*, 61(4):1338–1361 (electronic), 2000/01.

[31] T. F. Chan, S. H. Kang, and J. Shen. Total variation denoising and enhancement of color images based on the CB and HSV color models. *J. Visual Comm. Image Rep.*, 12(4):422–435, 2001.

[32] T. F. Chan, S. H. Kang, and J. Shen. Euler's elastica and curvature-based inpainting. *SIAM J. Appl. Math.*, 63(2):564–592 (electronic), 2002.

[33] T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043 (electronic), 2001/02.

[34] T. F. Chan and J. Shen. *Image processing and analysis*. SIAM, Philadelphia, PA, 2005.

[35] T. F. Chan and J. Shen. Variational image inpainting. *Comm. Pure Appl. Math.*, 58(5):579–619, 2005.

[36] P. Clément. Approximation by finite element functions using local regularization. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. RAIRO Analyse Numérique*, 9(R-2):77–84, 1975.

[37] R. Cohen, R. Hardt, D. Kinderlehrer, S. Y. Lin, and M. Luskin. Minimum energy configurations for liquid crystals: computational results. In *Theory and applications of liquid crystals (Minneapolis, Minn., 1985)*, volume 5 of *IMA Vol. Math. Appl.*, pages 99–121. Springer, New York, 1987.

[38] R. Cohen, S. Y. Lin, and M. Luskin. Relaxation and gradient methods for molecular orientation in liquid crystals. *Comput. Phys. Comm.*, 53(1-3):455–465, 1989. Practical iterative methods for large scale computations (Minneapolis, MN, 1988).

[39] G. Cortesani. Strong approximation of GSBV functions by piecewise smooth functions. *Ann. Univ. Ferrara Sez. VII (N.S.)*, 43:27–49 (1998), 1997.

[40] G. Cortesani. A finite element approximation of an image segmentation problem. *Math. Models Methods Appl. Sci.*, 9(2):243–259, 1999.

[41] G. Cortesani and R. Toader. A density result in SBV with respect to non-isotropic energies. *Nonlinear Anal.*, 38(5, Ser. B: Real World Appl.):585–604, 1999.

[42] G. Dal Maso. *An introduction to* Γ*-convergence*. Progress in Nonlinear Differential Equations and their Applications, 8. Birkhäuser Boston Inc., Boston, MA, 1993.

[43] E. De Giorgi. Free discontinuity problems in calculus of variations. In *Frontiers in pure and applied mathematics*, pages 55–62. North-Holland, Amsterdam, 1991.

[44] E. De Giorgi. Some remarks on Γ-convergence and least squares method. In *Composite media and homogenization theory (Trieste, 1990)*, volume 5 of *Progr. Nonlinear Differential Equations Appl.*, pages 135–142. Birkhäuser Boston, Boston, MA, 1991.

[45] E. De Giorgi. Variational free-discontinuity problems. In *International Conference in Memory of Vito Volterra (Italian) (Rome, 1990)*, volume 92 of *Atti Convegni Lincei*, pages 133–150. Accad. Naz. Lincei, Rome, 1992.

[46] E. De Giorgi and L. Ambrosio. New functionals in the calculus of variations. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 82(2):199–210 (1989), 1988.

[47] E. De Giorgi, M. Carriero, and A. Leaci. Existence theorem for a minimum problem with free discontinuity set. *Arch. Rational Mech. Anal.*, 108(3):195–218, 1989.

[48] K. Deckelnick and G. Dziuk. Error analysis of a finite element method for the Willmore flow of graphs. *Interfaces Free Bound.*, 8(1):21–46, 2006.

[49] F. Dibos and E. Séré. An approximation result for the minimizers of the Mumford-Shah functional. *Boll. Un. Mat. Ital. A (7)*, 11(1):149–162, 1997.

[50] M. Droske and A. Bertozzi. Higher-order feature-preserving geometric regularization. Preprint, URL: http://www.math.ucla.edu/ bertozzi/publications.html, 2009.

[51] M. Droske and M. Rumpf. A level set formulation for Willmore flow. *Interfaces Free Bound.*, 6(3):361–378, 2004.

[52] Q. Du, M. D. Gunzburger, and J. S. Peterson. Analysis and approximation of the Ginzburg-Landau model of superconductivity. *SIAM Rev.*, 34(1):54–81, 1992.

[53] Q. Du, C. Liu, and X. Wang. A phase field approach in the numerical study of the elastic bending energy for vesicle membranes. *J. Comput. Phys.*, 198(2):450–468, 2004.

[54] Q. Du and X. Wang. Convergence of numerical approximations to a phase field bending elasticity model of membrane deformations. *Int. J. Numer. Anal. Model.*, 4(3-4):441–459, 2007.

[55] G. Dziuk. Computational parametric Willmore flow. *Numer. Math.*, 111(1):55–80, 2008.

[56] G. Dziuk, E. Kuwert, and R. Schätzle. Evolution of elastic curves in $\mathbb{R}^n$: existence and computation. *SIAM J. Math. Anal.*, 33(5):1228–1245 (electronic), 2002.

[57] S. Esedoglu and J. Shen. Digital inpainting based on the Mumford-Shah-Euler image model. *European J. Appl. Math.*, 13(4):353–370, 2002.

[58] L. C. Evans. Partial regularity for stationary harmonic maps into spheres. *Arch. Rational Mech. Anal.*, 116(2):101–113, 1991.

[59] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[60] M. D. Fairchild. *Color Appearance Models*. Wiley-IS&T. Wiley, Chichester, West Sussex, second edition, 2006.

[61] H. Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[62] M. Focardi. On the variational approximation of free-discontinuity problems in the vectorial case. *Math. Models Methods Appl. Sci.*, 11(4):663–684, 2001.

[63] M. Focardi. *Variational Approximation of Vectorial Free Discontinuity Problems: the Discrete and Continuous Case*. PhD thesis, Scuola Normale Superiore di Pisa, 2002. URL: http://cvgmt.sns.it/papers/foc01/.

[64] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*, volume 224 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1983.

[65] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations*, volume 5 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1986. Theory and algorithms.

[66] E. Giusti. *Minimal surfaces and functions of bounded variation*, volume 80 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1984.

[67] J. Haehnle. Numerical approximations of the Mumford-Shah functional for unit vector fields. Preprint, URL: http://na.uni-tuebingen.de/preprints.shtml, 2009.

[68] R. M. Hardt. Singularities of harmonic maps. *Bull. Amer. Math. Soc. (N.S.)*, 34(1):15–34, 1997.

[69] F. Hélein. Régularité des applications faiblement harmoniques entre une surface et une variété riemannienne. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(8):591–596, 1991.

[70] J. G. Heywood and R. Rannacher. Finite element approximation of the nonstationary Navier-Stokes problem. I. Regularity of solutions and second-order error estimates for spatial discretization. *SIAM J. Numer. Anal.*, 19(2):275–311, 1982.

[71] C. A. Izmailov and E. N. Sokolov. The conceptual reflex arc: A model of neural processing as developed for colour vision. In H.-G. Geissler, editor, *Modern Issues in Perception*, pages 192–216. North-Holland, 1983.

[72] S. H. Kang and R. March. Variational models for image colorization via chromaticity and brightness decomposition. *IEEE Trans. Image Process.*, 16(9):2251–2261, 2007.

[73] B. Kawohl. From Mumford-Shah to Perona-Malik in image processing. *Math. Methods Appl. Sci.*, 27(15):1803–1814, 2004.

[74] R. Kimmel and N. Sochen. Orientation diffusion or how to comb a porcupine? *Journal of Visual Communication and Image Representation*, 13:238–248, 2002.

[75] F.-H. Lin. Nonlinear theory of defects in nematic liquid crystals; phase transition and flow phenomena. *Comm. Pure Appl. Math.*, 42(6):789–814, 1989.

[76] F.-H. Lin. Gradient estimates and blow-up analysis for stationary harmonic maps. *Ann. of Math. (2)*, 149(3):785–829, 1999.

[77] F. H. Lin and C. Y. Wang. Stable stationary harmonic maps to spheres. *Acta Math. Sin. (Engl. Ser.)*, 22(2):319–330, 2006.

[78] S. Y. Lin and M. Luskin. Relaxation methods for liquid crystal problems. *SIAM J. Numer. Anal.*, 26(6):1310–1324, 1989.

[79] C. Liu and N. J. Walkington. Approximation of liquid crystal flows. *SIAM J. Numer. Anal.*, 37(3):725–741 (electronic), 2000.

[80] C. Liu and N. J. Walkington. Mixed methods for the approximation of liquid crystal flows. *M2AN Math. Model. Numer. Anal.*, 36(2):205–222, 2002.

[81] P. Loreti and R. March. Propagation of fronts in a nonlinear fourth order equation. *European J. Appl. Math.*, 11(2):203–213, 2000.

[82] R. March and M. Dozio. A variational method for the recovery of smooth boundaries. *Image Vision Comput.*, 15(9):705–712, 1997.

[83] S. Masnou and J.-M. Morel. Level lines based disocclusion. In *Proc. IEEE ICIP'98*, volume 3, pages 259–263, 1998.

[84] L. Modica and S. Mortola. Un esempio di Γ^--convergenza. *Boll. Un. Mat. Ital. B (5)*, 14(1):285–299, 1977.

[85] D. Mumford. Elastica and computer vision. In *Algebraic geometry and its applications (West Lafayette, IN, 1990)*, pages 491–506. Springer, New York, 1994.

[86] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42(5):577–685, 1989.

[87] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, segmentation and depth*, volume 662 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1993.

[88] S. J. Osher and L. A. Vese. Numerical methods for p-harmonic flows and applications to image processing. *SIAM J. Numer. Anal.*, 40(6):2085–2104 (electronic) (2003), 2002.

[89] P. Perona and A. W. O. Diffusions. Orientation diffusions. *IEEE Trans. Image Processing*, 7:457–467, 1999.

[90] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.

[91] A. Prohl. *Computational micromagnetism*. Advances in Numerical Mathematics. B. G. Teubner, Stuttgart, 2001.

[92] M. Röger and R. Schätzle. On a modified conjecture of De Giorgi. *Math. Z.*, 254(4):675–714, 2006.

[93] J. C. Russ. *The Image Processing Handbook*. CRC Press, Boca Raton, FL, fifth edition, 2007.

[94] R. Schoen and K. Uhlenbeck. A regularity theory for harmonic maps. *J. Differential Geom.*, 17(2):307–335, 1982.

[95] R. Schoen and K. Uhlenbeck. Regularity of minimizing harmonic maps into the sphere. *Invent. Math.*, 78(1):89–100, 1984.

[96] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *IEEE Trans. Image Process.*, 7(3):310–318, 1998.

[97] M. Struwe. On the evolution of harmonic mappings of Riemannian surfaces. *Comment. Math. Helv.*, 60(4):558–581, 1985.

[98] M. Struwe. Geometric evolution problems. In *Nonlinear partial differential equations in differential geometry (Park City, UT, 1992)*, volume 2 of *IAS/Park City Math. Ser.*, pages 257–339. Amer. Math. Soc., Providence, RI, 1996.

[99] M. Struwe. *Variational methods*, volume 34 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer-Verlag, Berlin, third edition, 2000. Applications to nonlinear partial differential equations and Hamiltonian systems.

[100] B. Tang, G. Sapiro, and V. Caselles. Diffusion of general data on non-flat manifolds via harmonic maps theory: The direction diffusion case. *Journal Computer Vision*, 36:149–161, 2000.

[101] B. Tang, G. R. Sapiro, and V. Caselles. Color image enhancement via chromaticity diffusion. *IEEE Transactions on Image Processing*, 10(5):701–707, May 2001.

[102] E. G. Virga. *Variational theories for liquid crystals*, volume 8 of *Applied Mathematics and Mathematical Computation*. Chapman & Hall, London, 1994.