Stefan Fürtinger

An Approach to Computing Binary Edge Maps for the Purpose of Registering Intensity Modulated Images

Dissertation

Doctoral Advisor:

Ao. Univ. Prof. Mag. Dr. Stephen Keeling

External Advisor:

Prof. Dr. Jan Modersitzki

Institute for Mathematics and Scientific Computing
Faculty of Natural Sciences
Karl–Franzens University Graz

November 2012

Acknowledgements

The most exciting phrase to hear in science, the one that heralds the most discoveries, is not "Eureka!" (I found it!) but "That's funny..."
Isaac Asimov

I wish to express my deepest gratitude to all people who were involved in the development of this dissertation. First and foremost I want to thank my doctoral advisor Prof. Stephen Keeling. Over the past three years I did not only get to know him on a professional but also on a personal level. Through countless extra hours of discussing, brainstorming and conferring he always kept his indomitable sense of humor. He became a trusted mentor helping me with invaluable advice. His constant support, guidance and unshakable enthusiasm enabled me to achieve more than I ever thought possible. For this my mere expression of thanks does not suffice.

I owe sincere thankfulness to my external advisor Prof. Jan Modersitzki from the Institute of Mathematics and Image Computing at the University of Lübeck. He did not only give me the chance to present my work in Lübeck but also reviewed this thesis with great care and provided many helpful suggestions.

I gratefully acknowledge the financial support of the special research center "Mathematical Optimization and Applications in Biomedical Sciences" funded by the Austrian Science Fund FWF that made the realization of this research work possible.

Finally, I wish to express my deepest gratitude to my great family and friends. A very special thanks goes to my wife Doris for her ongoing support, patience and sympathy during all lows and highs of the past three years. Her love gave me the strength to pursue my endeavors. I would not have been able to write this theses without her. Further, I want to sincerely thank my sister Stefanie for her love and unwavering belief in my potential. Also, my jovial thanks go to my brother in spirit Florian who has been a faithful friend and companion for over twenty years now. Finally, I am deeply obliged to my mother. She raised me to be who I am today. She always believed in me, encouraged me and helped me when I lost confidence. For this a mere thank you is not enough.

Graz, November 2012   Stefan Fürtinger

Contents

List of Figures

List of Figures

List of Algorithms

List of Algorithms

Introduction

Image processing has become an increasingly important scientific field over the last two decades. The technical progress in image acquisition devices has entailed a growing demand for specialized sophisticated methods to postprocess recorded visual data. One example is the fusion of multiple images from various sources to obtain a comprehensive view, e.g., combining magnetic resonance and computerized tomography data to aid in surgery [FHM00, Chap. 8]. Another example is the extraction of certain features by identifying like objects in images, e.g., comparing patient data to an idealized atlas to identify tissue groups [FHM00, Chap. 17]. The methods used are as numerous as the applications in practice. May it be medicine, astronomy, biology, robotics, video games or weather forecasts, image processing nowadays is a vast area that makes use of a wide range of modern techniques in mathematics, including functional analysis, differential geometry and statistics.

This work approaches the problems of image segmentation and image registration against the background of biomedical applications. The starting point of the research work that led to this dissertation was an application coming from biophysics. The objective was to map a Purkinje fiber network in the endocardium of one heart to the endocardium of another heart [FKPP11]. The edge maps arising were mapped one to the other by representing these zero area sets as diffuse images which have positive measure supports. An elastic registration approach was developed and successfully applied to the images thus obtained. The convincing results achieved by this strategy raised the idea of developing an extension to general edge maps. The goal was to design a method able to register images that manifest considerable intensity variations. For example images from a dynamic contrast enhanced magnetic resonance image sequence (DCE-MRIs) taken before, during and after the appearance of contrast agent have this property. If binary edge maps of such images were available, then by transforming these edges to diffuse surfaces, the approach developed to register diffuse images could be applied.

Thus, the second part of the research consisted in designing a novel approach for computing binary edge maps. Inspired by state-of-the-art segmentation techniques, a framework of cost functionals revolving around characteristic functions, i.e., edge maps, was designed. Based on variational calculus the associated optimality formulations were established. In contrast to existing methods, the edges obtained were truly binary, i.e., no thresholding had to be performed. This dissertation is a comprehensive account of the research work described here, carried out over a three-year period.

Chapter 2 presents an introduction to image segmentation and discusses modern and established techniques relevant in the context of this work. The review of existing methods is precise in formulation but brief in presentation and is purely intended to provide a compact overview of mathematical tools involved without being excessively dense in theoretical details. In Chapter 3 the proposed approach for the computation of binary edge

maps is introduced by a one-dimensional example. The presentation given focuses solely on modeling considerations, while deliberately omitting many mathematical details for the sake of a clear presentation of ideas. Chapter 4 then fills these gaps and presents a rigorous analysis of the introduced approach. We prove existence and uniqueness of minimizers, show that the proposed iteration has a fixed point and establish continuity of computed results with respect to initial conditions. A finite element discretization of the introduced approach is discussed in Chapter 5. We show existence and uniqueness of solutions in the discrete setting and prove their convergence to the respective quantities in the continuum for successively smaller step sizes. Further, the behavior of the approach in the limit of vanishing regularization is discussed. We prove consistency for vanishingly small parameter choice not only in the continuum but also in the discrete setting. Chapter 6 presents an extensive investigation of practical results obtained by the proposed approach and discusses its performance compared to existing segmentation techniques.

Chapter 7 briefly introduces the concept of image registration and derives the similarity measure and regularization term used in the proposed cost functional. By means of a one-dimensional example the developed registration strategy for edge maps based on successive deblurring is explained. Previously published theoretical results are summarized in Chapter 8. Existence proofs covering minimizers of the introduced registration functional as well as asymptotic considerations in the limit of vanishing deblurring are discussed. Furthermore, optimality conditions and a solution scheme are established. Chapter 9 introduces the discrete setting of the registration approach and addresses algorithmic details. Finally, Chapter 10 presents results obtained by embedding the introduced approach for the computation of binary edge maps in the discussed registration framework. It is shown that a combination of the developed strategies can indeed successfully register image sequences that are challenging for classical intensity based registration techniques. The notation used throughout this work is summarized in Appendix A. Classical results that are used frequently are listed in Appendix B. Finally, Appendix C presents some items that have been left out in the main part of the text mostly for the sake of readability.

<div style="text-align: right; font-size: 3em; font-weight: bold; color: gray;">2</div>

The Image Segmentation Problem

One of the most challenging problems in image processing is ironically one of the easiest for the human eye: image segmentation. Ever since techniques from computer vision and machine learning have been used to identify objects in still images and videos it became quite clear that the notion of image segmentation is ambiguous [AK06, Chap. 4]. Depending on the given image the human eye typically segments it either based on objects or homogeneous regions. Figure 2.1 illustrates these two notions of image segments. This observation motivates the two general strategies in image segmentation. One approach uses application specific considerations to model a simplified approximation to the raw image. This simpler image is then separated by sharp edges. The second strategy employs a priori information about the problem to determine and extract significant contours out of the raw image. Although these two approaches follow different ideas the striking similarity in each case is the importance of edges. Thus image segmentation and edge detection are closely intertwined.

In this work both strategies will be combined: in the following we will derive a model to compute a smooth approximation I_s of a given image $\tilde{I}$. Based on this approximation a segmentation of an associated *fuzzy* edge map $\tilde{E}$ will be used to obtain sharp *binary* edges for the raw image $\tilde{I}$.

2.1 Definition of an Image

In mathematics images are typically modeled as functions defined on some bounded domain $\Omega \subset \mathbb{R}^2$ mapping to $\mathbb{R}^n$ with $n \geq 1$. More specifically let $\Omega := (0,1)^2$ be the image domain, $I \in L^1(\Omega)$ and define the *total variation* of I by

$$\int_\Omega |DI| := \sup\left\{ \int_\Omega I \nabla \cdot \varphi \, d\boldsymbol{x} \,\Big|\, \varphi \in C_0^1(\Omega) : \|\varphi\|_{L^\infty(\Omega)} \leq 1 \right\},$$

and introduce the space of functions of bounded variations (see, e.g., [AK06, Sec. 2.2.2])

$$BV(\Omega) := \left\{ I \in L^1(\Omega) \,\Big|\, \int_\Omega |DI| < \infty \right\}.$$

We consider images to be elements of $BV(\Omega)$. Note that $I \in BV(\Omega)$ may exhibit discontinuities. Thus in contrast to classical Sobolev spaces the gradient of I is not understood as a function but rather as a measure [AK06, Sec. 2.2]. This observation will play an important role later. Though being an object in the continuum of functions this model is rooted in *digital imaging*. A digital image is a discrete quantity defined on a grid. The cells of this grid are usually called *pixels*, the size of the grid is the *image resolution*. Around the year 2010 images from an average digital consumer camera had a resolution of about

<div style="text-align: right;">3</div>

Figure 2.1: Two notions of image segmentation. An image segmented by a human test subject (left), part of the *Berkeley Segmentation Dataset* (a collection of $12\,000$ hand-labeled segmentations of $1\,000$ Corel dataset images from 30 human subjects) [MFTM01]. Edges computed by the method introduced in this work (right).

ten *Megapixels* which usually corresponds to the grid size 3648×2736, i.e., $9\,980\,928$ pixels total. The value assigned to each pixel depends on the kind of image at hand. One of the simplest ways to store a color image, for instance, is to assign a three-dimensional vector representing the channels red, green and blue to each pixel. More complex formats take a fourth channel, sometimes called alpha channel, allowing for transparency effects. For the sake of simplicity only *gray-scale* images are considered in this work. Thus the gray-scale intensity value of the image at each pixel is expressed not in terms of a vector but a number. An eight bit gray scale image, for instance, uses unsigned bytes, i.e., numbers between 0 (black) and 255 (white), to save each pixel's intensity value. Thus $2^8 = 256$ shades of gray can be displayed. In terms of the continuum of the associated mathematical models this means that we only consider real scalar valued functions. Figure 2.2 shows an artificial image represented as a gray scale map and as surface plot of a function.
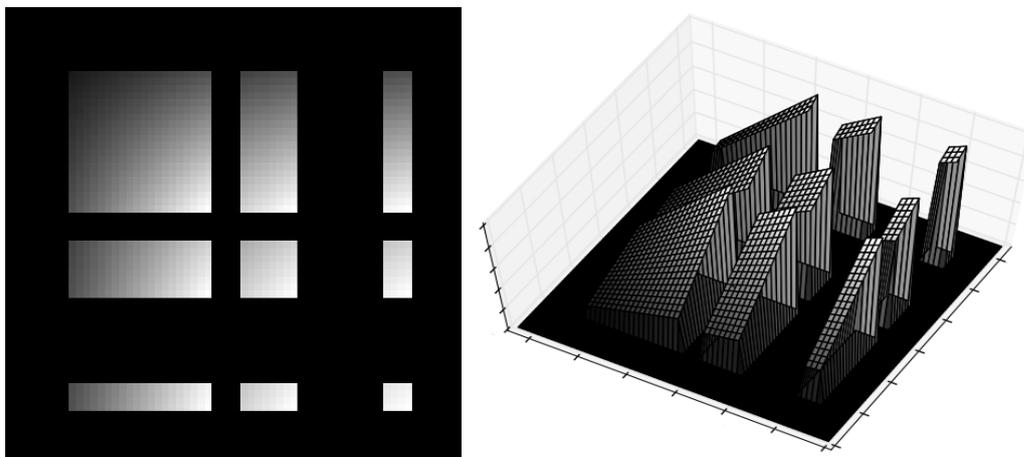
Figure 2.2: An artificial image represented as a gray scale map (left) and as a surface (right).
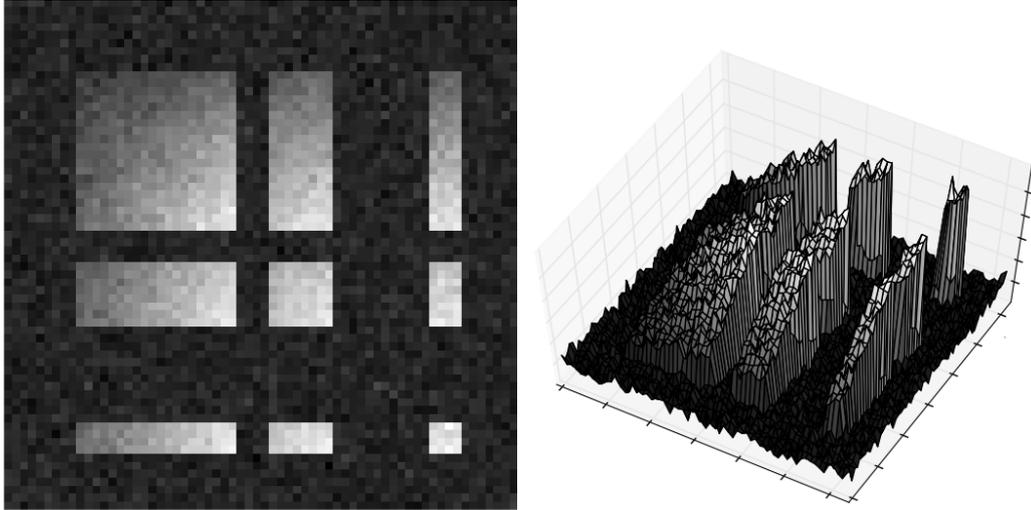
Figure 2.3: Influence of noise on an image. Shown is an artificial image corrupted by 5% additive Gaussian white noise (left) and the associated surface plot representation (right).

In the following we will always refer to images as functions of bounded variation mapping from $\Omega \subset \mathbb{R}^2$ to $\mathbb{R}$ unless explicitly stated otherwise.

2.2 Definition of an Edge

With the connection of segments and edges the first step in designing a segmentation approach is to establish a notion of edges appropriate for the task at hand. The biomedical applications of this work lend themselves better to the kind of edges seen in the right panel of Figure 2.1. Thus recall the surface plot depicted in Figure 2.2. Note that edges of the image shown correspond to jumps in the surface plot. We turn this heuristic observation into a more formal definition. The gradient of an image $I \in BV(\Omega)$ can be identified with a vector valued Radon measure that allows a decomposition into an absolutely continuous part, a jump part and a Cantor part [AK06, Sec. 2.2.3]. The jump part can be characterized by a corresponding jump set $S_I \subset \Omega$. We identify edges of an image with the jump set of its gradient. In the continuum this definition corresponds to the intuitive notion of edges. However, its direct application to the discrete setting is problematic. Since digital images are piecewise constant functions defined on a grid, any pixel whose intensity value is not exactly equal to those of its neighbors gives rise to an edge at the pixel's corresponding boundary. Hence in the discrete setting one usually relies on a more heuristic interpretation of jumps in an image. Typically an edge is thus considered to be a collection of points where the magnitude of the gradient $\nabla \tilde{I}$ is "large". The following investigation of this particular understanding of an edge is guided by the review in [CS05, Sec. 7.2.1]. In this edge paradigm the simplest edge detector is given by determining the set

$$\Gamma(\tau) := \left\{ \boldsymbol{x} \in \Omega \,:\, |\nabla \tilde{I}(\boldsymbol{x})| \geq \tau \right\},$$

for some threshold $\tau > 0$. Obviously, this formulation has some severe shortcomings. First and most significant is its sensitivity to *scaling*. Since the parameter τ is chosen entirely independent from the image $\tilde{I}$ any multiple $\kappa \tilde{I}$ with $\kappa \in \mathbb{R}$ yields a potentially different
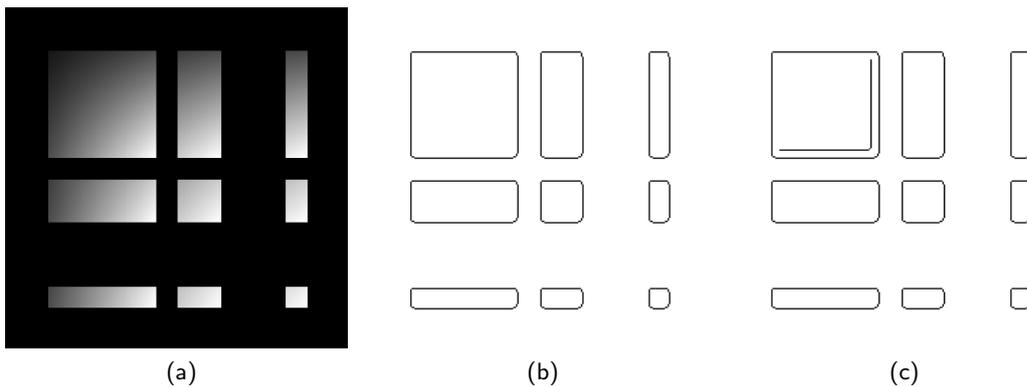
Figure 2.4: Result of Canny's Edge Detector applied to the artificial image (a) for $\tau = 0.05$ (b) and $\tau = 0.04$ (c).

edge set $\Gamma(\tau)$. However, scaling $\tilde{I}$ (and thus $\nabla \tilde{I}$) by κ has no influence on the formation of the image's edges. For similar reasons $\Gamma(\tau)$ is also very prone to noise in $\tilde{I}$. Consider the noise corrupted artificial image shown in Figure 2.3. The human eye easily recognizes that the noisy image has the same edges as the noise free version shown in Figure 2.2. Observe, however, that $\tilde{I}$ exhibits many scattered jumps, i.e., isolated local maxima of $\nabla \tilde{I}$, illustrated in the right plot of Figure 2.3. These disconnected points of significantly different intensity make a hard thresholding approach quite difficult and may thus severely impair $\Gamma(\tau)$.

A very popular remedy for these problems is as follows. Let $\boldsymbol{g}_{\mathrm{n}} := \nabla \tilde{I}/|\nabla \tilde{I}|$ denote the normalized gradient of $\tilde{I}$. Instead of working with (the possibly noisy) raw image $\tilde{I}$ itself consider a smooth approximation I_σ of $\tilde{I}$. Thus let I_σ be the convolution of $\tilde{I}$ with a Gaussian kernel. Then

$$\Gamma_{\mathrm{CE}}(\tau) := \left\{ \boldsymbol{x} \in \Omega \,\middle|\, \max_{\boldsymbol{g}_{\mathrm{n}}} |\nabla I_\sigma| \wedge |\nabla I_\sigma| \geq \tau \right\},$$

is the famous *Canny Edge Detector* [Can83] introduced in 1983. The remarkable idea behind $\Gamma_{\mathrm{CE}}(\tau)$ is to locally maximize $|\nabla I_\sigma|$ along $\boldsymbol{g}_{\mathrm{n}}$. Thus the set $\Gamma_{\mathrm{CE}}(\tau)$ consists of points at which the gradient magnitude is maximal in the gradient direction and not just "large" compared to some hard threshold. This restriction together with the usage of the mollification I_σ instead of $\tilde{I}$ makes it much less likely that noise corruption, i.e., isolated large gradient magnitudes, leads to erroneously detected edge points. However, even more importantly, by tracing edges along gradient vectors and not only relying on large gradients, $\Gamma_{\mathrm{CE}}(\tau)$ becomes insensitive to scaling.

Nevertheless, practical implementations of Canny's edge detector still crucially rely on well performed thresholding (compare, e.g., [RE95]). Furthermore, while being much less sensitive to changes in τ than the naive detector $\Gamma(\tau)$ edge sets $\Gamma_{\mathrm{CE}}(\tau)$ may vary significantly for different threshold values. Figure 2.4 depicts this behavior. Thus, albeit being one of the most widely used edge detectors nowadays, the biggest drawback of $\Gamma_{\mathrm{CE}}(\tau)$ is still its dependence on the threshold τ. Hence one of the objectives of this work is to develop a segmentation algorithm that does not rely on any thresholding techniques. However, considering a smooth approximation rather than the raw image itself is a technique that prevailed in image segmentation and is also utilized in this work.
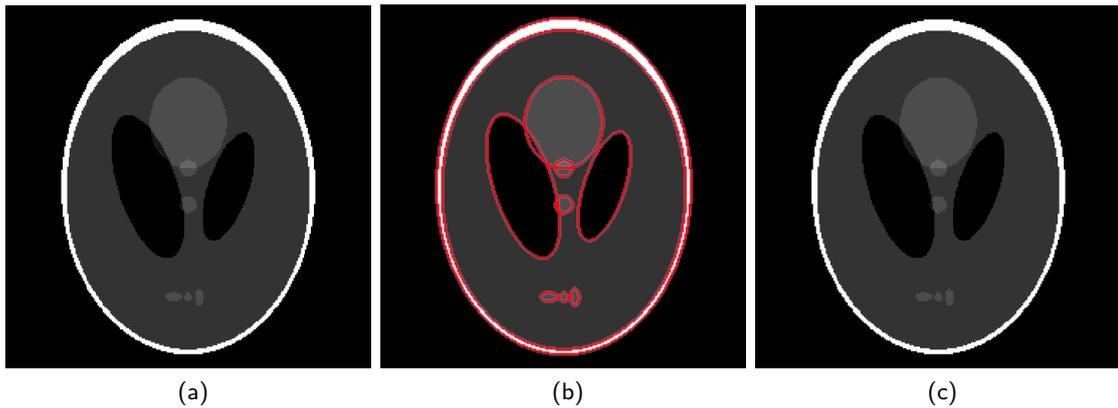
Figure 2.5: K-means segmentation of the Shepp–Logan phantom [SL74]. Panel (a) shows the raw image. Note that the image is piecewise constant. Panel (b) shows the edges of the K-means segments superimposed on the raw image. Panel (c) depicts the K-means approximation $I_{K\mathrm{M}}$.

2.3 Approximating the Raw Image

One of the simplest ways to approximate a given image $\tilde{I}$ such that sharp edges are obtained is the method of *K-means clustering*. We denote by $\{\Omega_k\}_{k=1}^{K}$ a disjoint partitioning of the image domain Ω, i.e., $\Omega = \dot{\cup}_{k=1}^{K}\Omega_k$, which means $\Omega = \cup_{k=1}^{K}\Omega_k$ such that $\Omega_i \cap \Omega_j = \emptyset$ if $i \neq j$. The approach of K-means clustering constructs an approximation $I_{K\mathrm{M}}$ of $\tilde{I}$ such that $I_{K\mathrm{M}} = I_k$ on Ω_k for some $I_k \in \mathbb{R}$. Thus $I_{K\mathrm{M}} = \sum_{k=1}^{K} I_k \chi_k$ where χ_k denotes the characteristic function of Ω_k. Formally $I_{K\mathrm{M}}$ is a solution of

$$\min_{\{I_k\},\{\chi_k\}} \int_{\Omega} \left| \sum_{k=1}^{K} I_k \chi_k - \tilde{I} \right|^2 d\boldsymbol{x} \text{ s.t. } \Omega = \dot{\cup}_{k=1}^{K}\Omega_k.$$

Note, however, that in practice $I_{K\mathrm{M}}$ is usually computed by an iterative refinement algorithm (see, e.g., [Bis06, Sec. 9.1] for a detailed review). Based on its intensity value a pixel of $\tilde{I}$ is assigned to the segment Ω_k that has the closest mean value I_k. This procedure is often referred to as *Assignment Step*. After each pixel has been associated to a segment the mean intensity value I_k of each Ω_k is recalculated in the so-called *Update Step*. As soon as the segments Ω_k cease to change the algorithm terminates. Although the computational performance of this method is usually very good it is purely heuristic and cannot be guaranteed to converge to a global minimum in general [Bis06, p. 425]. Furthermore, the quality of the final result $I_{K\mathrm{M}}$ strongly depends on the initialization of segments and associated means. Figure 2.5 shows a K-means segmentation of some exemplary image. However, besides these practical problems, K-means clustering also has some serious theoretical shortcomings. First of all the computation of mean value clusters is based on the assumption of piecewise constancy of the raw image $\tilde{I}$. The segmentation performance of K-means severely degenerates for images violating this assumption or in the presence of noise. To illustrate this behavior, Figure 2.6 depicts a one-dimensional example. The raw data $\tilde{I}$ (shown in black) is a piecewise linear function on $(0, 1)$ but obviously not piecewise constant. A natural choice for the optimal number of segments for this example is two,
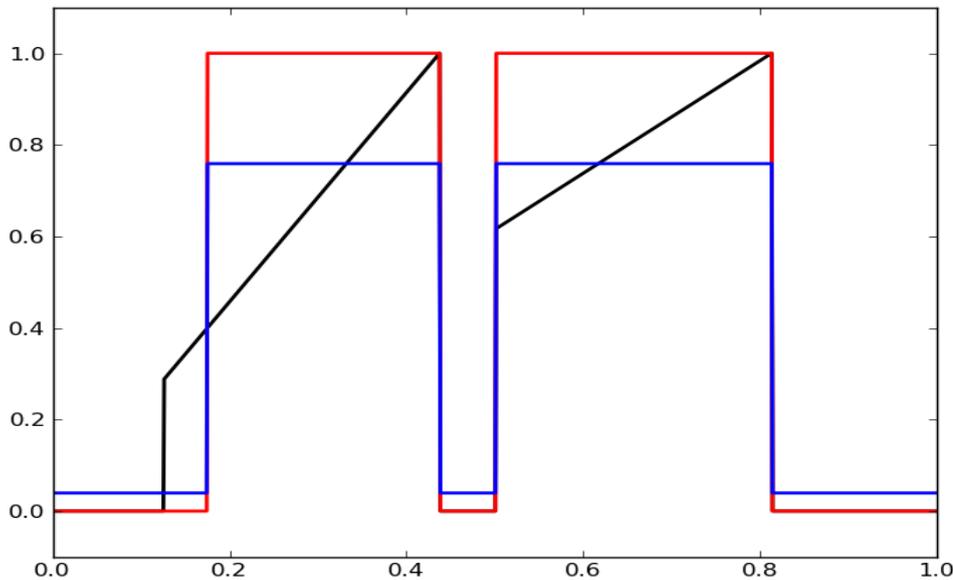
Figure 2.6: The importance of piecewise constancy for the performance of K-means: a 1D problem. Shown is $\tilde{I}$ (black), the approximation I_{KM} (blue) and the characteristic function χ (red) of one segment ($K = 2$).

such that $\Omega = \Omega_1 \cup \Omega_2$ with $\Omega_1 = \left\{ x \in \Omega \,\middle|\, \tilde{I}(x) = 0 \right\}$ and $\Omega_2 = \left\{ x \in \Omega \,\middle|\, \tilde{I}(x) \neq 0 \right\}$. Thus we set $K = 2$ to obtain the wanted segmentation that separates the two ramps seen in the plot such that

$$(2.1) \qquad I_{KM}|_{\Omega_1} = \frac{1}{|\Omega_1|} \int_{\Omega_1} \tilde{I}(x) \, dx \quad \text{and} \quad I_{KM}|_{\Omega_2} = \frac{1}{|\Omega_2|} \int_{\Omega_2} \tilde{I}(x) \, dx.$$

However, since $\tilde{I}|_{\Omega_2}$ is strictly monotone the method of K-means clustering results in an unwanted segmentation. The red line shows the characteristic function χ of one computed segment, i.e., $\chi = 0$ on the first segment and $\chi = 1$ otherwise. Though the right ramp of $\tilde{I}$ is segmented correctly the leftmost part of its other ramp has been erroneously cut out and assigned to Ω_1. Thus Ω_1 as computed by K-means contains nonzero regions of $\tilde{I}$. This makes the approximation I_{KM} (shown in blue) strictly positive on the whole of Ω and hence the expected mean value approximation (2.1) is not satisfied. Thus violating the assumption of piecewise constancy may indeed severely impair the outcome of K-means. Note that setting $K = 3$ is not a remedy here since this would introduce an unwanted third segment.

This leads to a further drawback of the K-means clustering approach: the number of segments K has to be specified a priori. Depending on the given image $\tilde{I}$ choosing K may be non-trivial: if K is too small, distinct objects may be lumped together in the same segment (see Figure 2.7 (b) and (c)). Conversely, if K is too large the algorithm tends to divide the image into an unnaturally large number of pieces, where all segments are necessarily disjoint but a given segment consists of many disconnected components (Panels (e) and (f) of Figure 2.7). Perhaps the most serious drawback of the K-means

approach is, however, that irregularity of segments is not penalized. Thus very complicated edges exhibiting, e.g., oscillatory behavior, may arise if the image $\tilde{I}$ is not smooth. This renders the algorithm almost useless for noisy data or images exhibiting oscillatory textures as can be seen in the bottom row panels of Figure 2.7. Though this may be remedied to some degree by preprocessing (discussed in the next section) the raw image $\tilde{I}$, regularity of segments generated by K-means cannot be imposed in general.
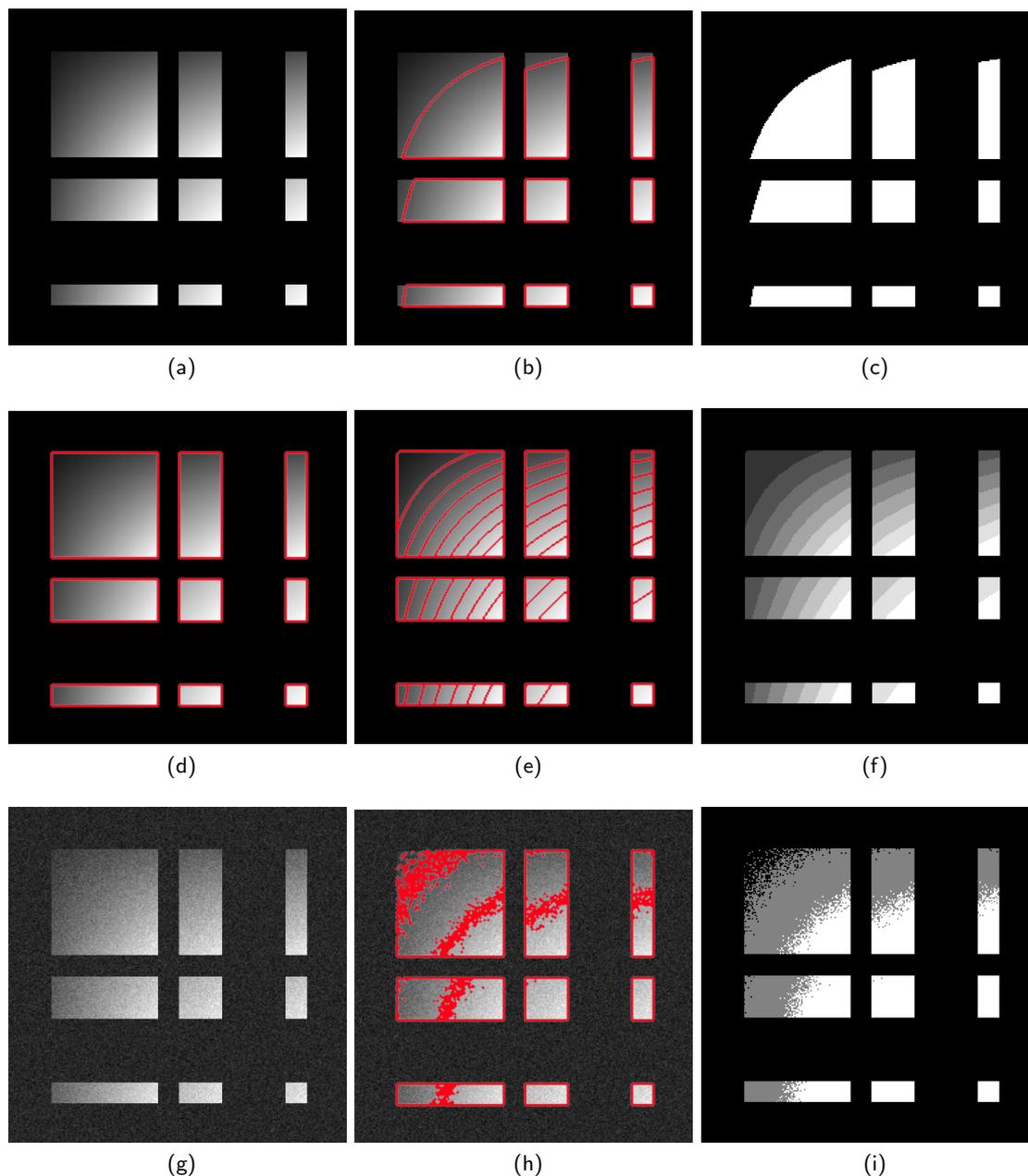
Figure 2.7: Performance of K-means on the (not piecewise constant) image depicted in Panel (a). A favorable segmentation is shown in Panel (d). The middle image in the top row shows the boundaries of segments computed by K-means clustering with $K = 2$ superimposed on the raw image. Panel (c) depicts the corresponding reconstruction I_{KM}. The same format is used in Panels (e) and (f) for the case $K = 9$ and in Panels (h) and (i) for the case $K = 3$. The bottom row illustrates the performance of K-means applied to the image (g), obtained by adding 5% Gaussian noise to the image in (a).

2.4 Edge Preserving Preprocessing

We have seen that noise corruption poses a substantial problem in the context of edge detection. A natural way to overcome difficulties arising from noisy input data is to
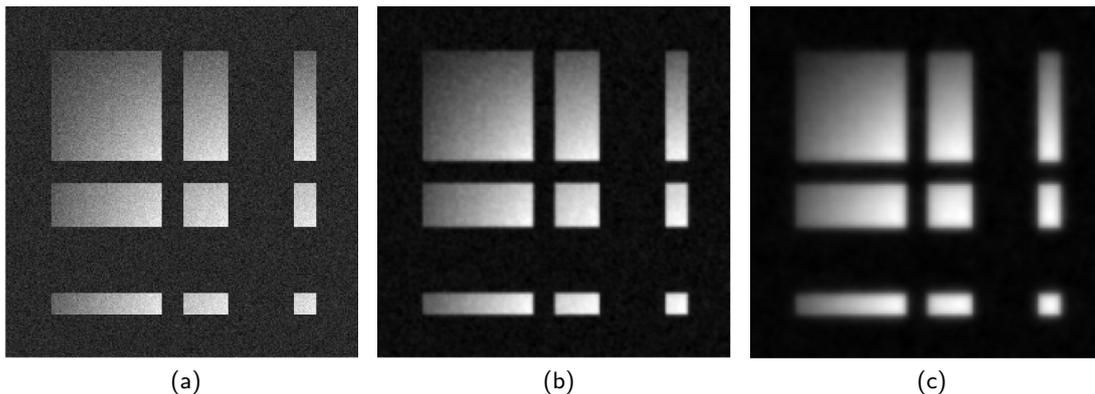
Figure 2.8: Denoising results obtained by minimizing J_{TA}. Shown is an artificial image corrupted by 5% additive Gaussian white noise (a) and a minimizer of J_{TA} for $\nu = 1e-5$ (b) and $\nu = 1e-4$ (c).

preprocess the raw image in a way that diminishes noise degradation. This processing is called *image restoration* or *image denoising* (see, e.g., [CS05, Chap. 4]). The first step to approaching this problem is to set up a mathematical framework that models noise corruption of an image. Thus assume I_{orig} is the unknown original image, $\tilde{I}$ is again the (measured) raw image and η is some noise term. One of the simplest models linking these three quantities is the following linear relation

$$(2.2) \qquad\qquad \tilde{I} = I_{\text{orig}} + \eta.$$

Since we know nothing about the noise η, recovering I_{orig} from $\tilde{I}$ by relying solely on the model (2.2) is a complex task. If η is white Gaussian noise then following the Maximum Likelihood Principle (compare [AK06, Sec. 3.2.1]), a first attempt to compute an approximation of I_{orig} is to solve the least squares problem

$$(2.3) \qquad\qquad \inf_{I} \int_{\Omega} \frac{1}{2} \left| \tilde{I} - I \right|^2 d\boldsymbol{x}.$$

Note that the global minimum of this problem is zero which is attained at $I = \tilde{I}$. Thus this approach is not feasible here. A quite natural remedy in this situation is provided by the concept of *regularization*. The idea is to not only impose data fidelity as in (2.3) but to also enforce wanted properties of a solution by introducing an appropriate penalty term [AK06, Sec. 3.2.2]. Since the objective is to remove noise from $\tilde{I}$ a reasonable demand on solutions is to have a low gradient (compare the discussion on isolated large gradient magnitudes in Section 2.2). Following this concept consider the optimization problem to minimize

$$J_{\text{TA}}[I] := \frac{1}{2} \int_{\Omega} \frac{1}{\nu} \left| \tilde{I} - I \right|^2 + |\nabla I|^2 \, d\boldsymbol{x} = \frac{1}{2\nu} \left\| \tilde{I} - I \right\|_{L^2(\Omega)}^2 + \frac{1}{2} \left\| \nabla I \right\|_{L^2(\Omega)}^2 ,$$

where $\nu > 0$ is a weighting parameter. A similar functional was first published in 1977 by Tikhonov and Arsenin [TA77]. The first term in J_{TA} is a so-called *data fidelity term* ensuring that the computed approximation I is "close" to the given image $\tilde{I}$. The second part of J_{TA} is the so-called *regularization term* that penalizes large gradients. Thus

11

potential minimizers of J_{TA} need not only to approximate $\tilde{I}$ but must also have minimal gradient magnitude. The minimization of J_{TA} is well studied; thus it has been shown (e.g., [AK06, Chap. 3]) that this optimization problem has a solution. For computing a minimizer of J_{TA} one usually derives necessary optimality conditions by means of *Gâteaux differentiation* of J_{TA} (see Appendix B.2 for details). This yields a *weak formulation*, i.e., a variational problem in some function space, that can be transferred to a *strong formulation*, i.e., a partial differential equation (PDE), the so-called *Euler–Lagrange equation* associated to the optimization problem. The Euler–Lagrange equation that arises from minimizing J_{TA} is given by

$$(2.4) \qquad \begin{cases} -\nu \Delta I + I = \tilde{I}, & \text{in } \Omega, \\ \dfrac{\partial}{\partial \boldsymbol{n}} I = 0, & \text{on } \partial\Omega, \end{cases}$$

where $\boldsymbol{n}$ denotes the outer unit normal vector on $\partial\Omega$. A detailed derivation is given in Appendix C.1. The Euler–Lagrange equation highlights the most significant drawback of the functional J_{TA}. The Laplacian appearing in (2.4) has strong isotropic smoothing properties [AK06, Sec. 3.2.2]. Thus denoising a image $\tilde{I}$ by minimizing J_{TA} may result in a blurry solution as illustrated in Figure 2.8. Thus, albeit clean of noise, minimizers of J_{TA} usually exhibit washed out edges since $\|\nabla I\|_{L^2(\Omega)}$ penalizes *all* large gradients not only those corresponding to unwanted isolated jumps. One of the most widely used remedies for this situation was introduced in 1992 by Rudin, Osher and Fatemi in their famous paper [ROF92]. They proposed penalizing the $L^1(\Omega)$-norm of ∇I instead of the $L^2(\Omega)$-norm resulting in the functional

$$(2.5) \qquad J_{\mathrm{ROF}}[I] = \frac{1}{2\nu} \left\| \tilde{I} - I \right\|_{L^2(\Omega)}^2 + \frac{1}{2} \|\nabla I\|_{L^1(\Omega)}.$$

Note that if I is sufficiently smooth, $\|\nabla I\|_{L^1(\Omega)}$ is the total variation of I which motivated the use of the term *TV-regularization*. However, the $L^1(\Omega)$-norm in (2.5) leads to a non-linear Euler–Lagrange equation (compare, e.g., [CL97]). A natural choice for a space to minimize (2.5) is $\{I \in L^2(\Omega) | \nabla I \in L^1(\Omega)\}$. However, this space does not lend itself handily for the investigation of minimizers of J_{ROF} (compare, e.g., the discussion in [AK06, Sec. 3.2.3]). Thus, one minimizes J_{ROF} in $BV(\Omega)$. Unfortunately the fundamental Lemma of Calculus of Variations cannot be applied directly to $BV(\Omega)$. Thus the Euler–Lagrange equations associated to minimizing J_{ROF} on $BV(\Omega)$ become even more complex (for a thorough investigation of the space of functions of bounded variations and its role in image restoration see, e.g., [CCN11], [CL97] and particularly the review in [AK06, Sec. 2.2, Sec. 3.2.3]). Moreover, the numerical solution of non-linear PDEs is in general computationally expensive. Thus there has been strong interest in the development of fast solution schemes for the minimization of J_{ROF}. To date the most widely used strategy is a projection algorithm introduced in 2004 by Chambolle [Cha04]. In the following we will consider Chambolle's Algorithm in greater detail since it does not only iteratively approximate the minimizer of J_{ROF} but as a by-product it also computes a smooth vector field associated to the edges of $\tilde{I}$. The following discussion loosely follows the presentation given in [CCN11] for the discrete setting.

Algorithm 2.1 Chambolle's projection approach for the TV denoising problem

---

    Input: $\tilde{I}$, ν
    Output: I_{TV}, $\boldsymbol{p}$

1: **Choose** $0 < \Delta t < \frac{1}{8}$ and set $I_{\mathrm{TV}} \leftarrow \tilde{I}$ as well as $\boldsymbol{p} \leftarrow 0$

2: **for** $k = 1, \dots$ **do**
3:     **Compute**
$$\boldsymbol{p} = \frac{\boldsymbol{p} + \Delta t \nabla \left(\nu \nabla \cdot \boldsymbol{p} - \tilde{I} \right)}{1 + \Delta t \left| \nabla \left(\nu \nabla \cdot \boldsymbol{p} - \tilde{I} \right) \right|}.$$
4:     **Update**
$$I_{\mathrm{TV}} = \tilde{I} + \nu \nabla \cdot \boldsymbol{p}$$
5: **end for**

Using notation found in, e.g., [GR05, p. 135], we introduce the set

$$\mathcal{K} := \left\{ v \in L^2(\Omega) \,|\, \exists \boldsymbol{p} \in H_0(\mathrm{div}) : |\boldsymbol{p}| \leq 1 \text{ a.e. } \Omega \text{ such that } v = \nabla \cdot \boldsymbol{p} \right\},$$

with

$$H_0(\mathrm{div}) := \left\{ \boldsymbol{p} \in L^2(\Omega) \,\middle|\, \nabla \cdot \boldsymbol{p} \in L^2(\Omega), \, \boldsymbol{p} \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega \right\}.$$

Then it can be shown [BL10, Sec. 1.3] by using techniques from convex analysis and Fenchel duality that the minimizer of J_{ROF} on $BV(\Omega)$ is given by

$$I_{\mathrm{TV}} = \tilde{I} - P_{\nu\mathcal{K}}[\tilde{I}],$$

where $P_{\nu\mathcal{K}}$ is the $L^2(\Omega)$-orthogonal projection of $\tilde{I}$ onto the set $\nu\mathcal{K}$. Thus the problem of minimizing J_{ROF} is shifted to computing the projection $P_{\nu\mathcal{K}}[\tilde{I}]$. The projection $P_{\nu\mathcal{K}}[\tilde{I}]$ on the other hand is given by the solution of the constrained minimization problem

$$\min_{\boldsymbol{p} \in H_0(\mathrm{div})} \left\| \nu \nabla \cdot \boldsymbol{p} - \tilde{I} \right\|_{L^2(\Omega)}^2 \quad \text{such that } |\boldsymbol{p}| \leq 1 \text{ a.e. } \Omega.$$

A standard technique to solve this constrained problem is to study the associated Lagrange function. It can be shown [Cha04] that there indeed exists a Lagrange multiplier α such that we obtain the Euler–Lagrange equation

$$-\nabla \left(\nu \nabla \cdot \boldsymbol{p} - \tilde{I} \right) + \alpha \boldsymbol{p} = 0,$$

where $\alpha = 0$ and $|\boldsymbol{p}| < 1$ or $\alpha = 1$ and $|\boldsymbol{p}| = 1$. Thus $\alpha = \left| \nabla \left(\nu \nabla \cdot \boldsymbol{p} - \tilde{I} \right) \right|$ satisfies this requirement which yields

$$(2.6) \qquad -\nabla \left(\nu \nabla \cdot \boldsymbol{p} - \tilde{I} \right) + \left| \nabla \left(\nu \nabla \cdot \boldsymbol{p} - \tilde{I} \right) \right| \boldsymbol{p} = 0.$$
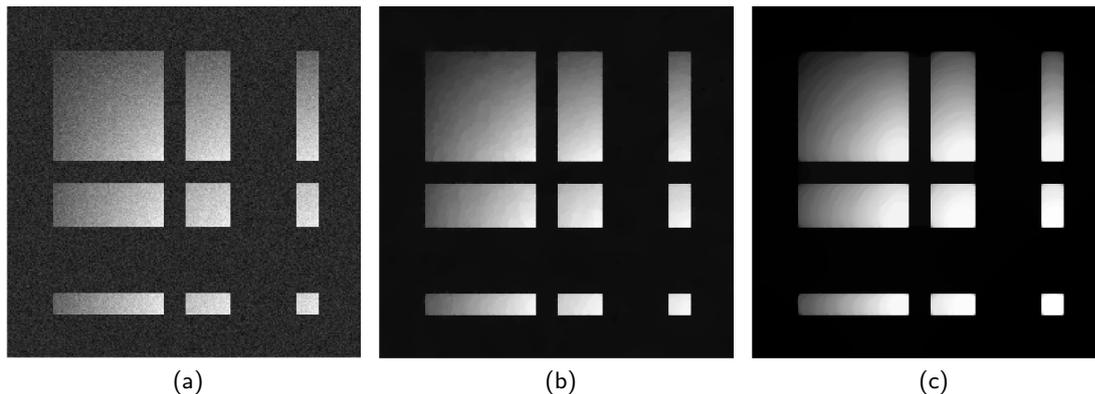
Figure 2.9: TV denoising results obtained by using Chambolle's projection approach (Algorithm 2.1). Shown is an artificial image corrupted by 5% additive Gaussian white noise (a) and the TV denoised image for $\nu = 0.05$ (b) and $\nu = 0.4$ (c). The Algorithm was run to numerical convergence.

As a solution technique one introduces a pseudo-time t and the term $-\frac{\partial}{\partial t}\boldsymbol{p}$ on the right hand side of (2.6). Using a semi explicit discretization with step size Δt this yields

$$\nabla\left(\nu\nabla\cdot\boldsymbol{p}^k - \tilde{I}\right) - \left|\nabla\left(\nu\nabla\cdot\boldsymbol{p}^k - \tilde{I}\right)\right|\boldsymbol{p}^{k+1} = \frac{\boldsymbol{p}^{k+1} - \boldsymbol{p}^k}{\Delta t},$$

or equivalently

$$\boldsymbol{p}^{k+1} = \frac{\boldsymbol{p}^k + \Delta t\nabla\left(\nu\nabla\cdot\boldsymbol{p}^k - \tilde{I}\right)}{1 + \Delta t\left|\nabla\left(\nu\nabla\cdot\boldsymbol{p}^k - \tilde{I}\right)\right|}.$$

Chambolle proved [Cha04] that in the discrete setting the divergence of $\boldsymbol{p}^k$ given by this iteration converges indeed to the projection of $\tilde{I}$ onto $\nu\mathcal{K}$ as $k \to \infty$ provided that $\Delta t < 1/8$. Thus we summarize Chambolle's projection approach in Algorithm 2.1. Observe that in contrast to the minimization of J_{TA} that involved solving the Euler–Lagrange equation (2.4), Algorithm 2.1 does not require the solution of any linear system and is thus a purely explicit approach. This makes Chambolle's projection strategy computationally very efficient compared to other numerical approaches.

Figure 2.9 illustrates the denoising capabilities of Chambolle's Algorithm for different values of ν. Note that in contrast to the results obtained by minimizing J_{TA} depicted in Figure 2.8 the TV-restored images exhibit sharp edges. Thus the use of $\|\nabla I\|_{L^1(\Omega)}$ in J_{ROF} instead of $\|\nabla I\|_{L^2(\Omega)}$ has indeed an edge-preserving effect. However, with a view to Figure 2.9 an unwanted phenomenon can be observed as well. In contrast to the clean uncorrupted image seen in Panel (a) of Figure 2.7 the reconstructions computed by minimizing J_{ROF} show some "blocky" artifacts in regions where the original image exhibits smooth intensity variations. This well known effect is often called *staircasing* [KS02]. Since the kernel of the TV regularizer comprises constant functions TV denoised images tend to be piecewise constant (for an extensive review of staircasing effects in image reconstruction see also [KS02]). Thus staircasing is not a peculiarity of Chambolle's projection algorithm but an inherent problem of TV regularization in general. A very elegant approach that augments TV regularization by enriching the regularizer's kernel is the
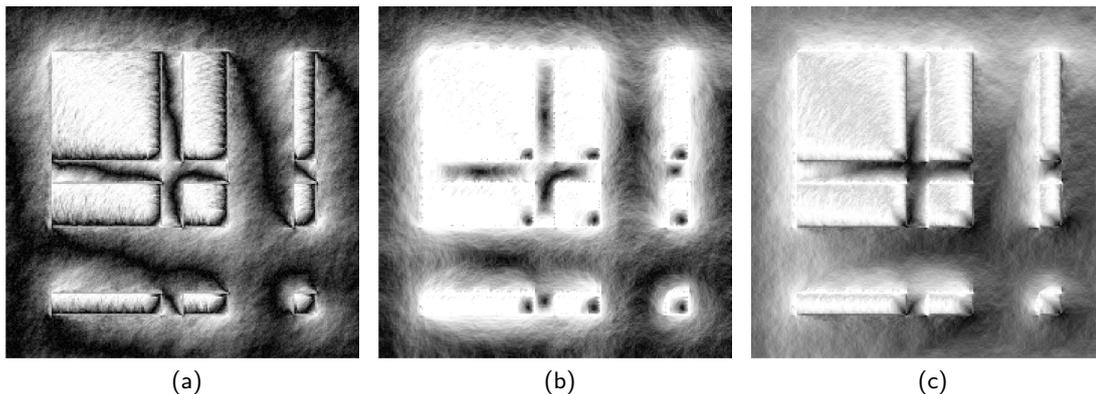
Figure 2.10: Visualization of the dual variable p from Chambolle's Algorithm 2.1. Shown is $|p|_1$ (a), $|p|_2$ (b) and $|p|_\infty$ (c) for $\nu = 0.4$.

so-called *Total Generalized Variation* [BKP10]. However, it is not within the scope of this work to analyze possible remedies for the staircasing phenomenon.

Instead, we are interested in the connection of edges and the dual variable p appearing in Chambolle's projection algorithm. Figure 2.10 shows $|p|_1$, $|p|_2$ and $|p|_\infty$ for $\nu = 0.4$ (the respective reconstruction I_{TV} is depicted in Panel (c) of Figure 2.9. Unsurprisingly the Euclidean norm (Panel (b) of Figure 2.10) yields the poorest results with respect to edge enhancement due to the averaging of components in every single point. Note, however, that edges of the original image are indeed amplified in $|p|_1$ and $|p|_\infty$. Thus the vector field p really can provide additional information regarding the image's edges. Hence $|p|_1$ and $|p|_\infty$ may be seen as some kind of *edge map* of the original image. However, both $|p|_1$ and $|p|_\infty$ are *fuzzy* which means neither $|p|_1$ nor $|p|_\infty$ provide a binary "yes or no" information whether a point is on an edge or not. Thus, if a boolean binary edge map is wanted some postprocessing of p is necessary. This ultimately requires some thresholding strategy and hence suffers from similar problems as, e.g., Canny's edge detector discussed in Section 2.2. Nevertheless, Chambolle edge maps are a rather popular tool in image processing not least because they emerge for free when employing Algorithm 2.1 in the context of image reconstruction problems.

2.5 Extracting Contours

A way to circumvent irregular segments that does not depend on preprocessing techniques was introduced in 1989 by Mumford and Shah in their seminal paper [MS89]. Let Γ denote a one-dimensional edge set. The *Mumford–Shah functional* is given by

$$J_{\mathrm{MS}}(I, \Gamma) = \frac{\alpha}{2} \int_\Omega \left| I - \tilde{I} \right|^2 d\boldsymbol{x} + \frac{\kappa}{2} \int_{\Omega \setminus \Gamma} |\nabla I|^2 d\boldsymbol{x} + \nu \mathcal{H}(\Gamma),$$

where $\mathcal{H}$ denotes the one-dimensional Hausdorff measure (see, e.g.,[Hal74, p. 53]). Here some important observations should be made. The first term in J_{MS} guarantees data fidelity. The second term has to be interpreted with reasonable care since Γ is a set of Lebesgue measure zero, i.e., $|\Gamma| = 0$. In the Mumford–Shah context I is explicitly assumed to be discontinuous on Γ so that ∇I is singular on Γ and we are thus interested
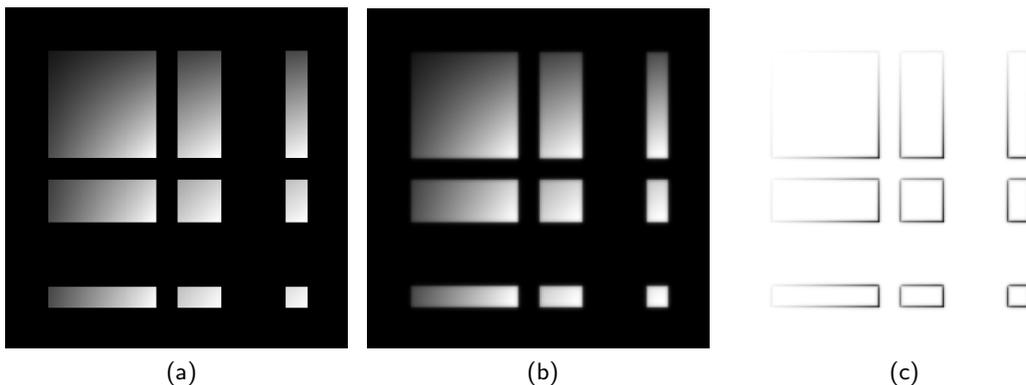
Figure 2.11: Exemplary Ambrosio–Tortorelli segmentation. Shown is the approximation (b) and phase function (c) for the image (a).

in functions $I \in H^1(\Omega \backslash \Gamma)$. Hence the second term gives an incentive for I to be smooth off edges but does not penalize "jumps" on the edge set Γ. Thus including Γ in the domain of integration permits the use of the $L^2(\Omega)$-norm of ∇I as regularization term on $\Omega \backslash \Gamma$ albeit avoiding the loss of edges (compare the discussion in the previous section). A motivation for the third term of J_{MS} can be given by looking again at the bottom panels of Figure 2.7: the oscillatory edges depicted in Panel (h) are in some sense "too long" compared to the desired optimal edges shown in Panel (d) of the same Figure. This motivates including $\mathcal{H}(\Gamma)$ in J_{MS} in order to penalize long edges.

A rigorous analysis of the Mumford–Shah functional still proves to be very difficult due to the appearance of the geometric quantity Γ as independent variable. Without further assumptions on the regularity of Γ not even existence of minimizers of J_{MS} can be guaranteed (compare, e.g., [AK06, Sec. 4.2]). Thus deriving optimality conditions for the Mumford–Shah functional is also only possible under stronger regularity assumptions. However, even then the discretization of the set of discontinuities Γ remains challenging. One of the most popular approaches to overcome these difficulties is to approximate J_{MS} by elliptic functionals defined solely on standard Sobolev spaces, thus eliminating the dependence on the geometric variable Γ. The idea behind this concept may be illustrated by the following heuristic considerations guided by the presentation given in [She05]. Instead of explicitly including an edge set in the cost functional, Γ is approximated by "diffuse regions" in the image. More specifically, the integral on $\Omega \backslash \Gamma$ and the length $\mathcal{H}(\Gamma)$ appearing in J_{MS} are replaced by local volume integrals of appropriate functions. This can be achieved by introducing a so-called *phase function* ψ_ε which might ideally be given by

$$(2.7) \qquad \psi_\varepsilon(\boldsymbol{x}) := \begin{cases} \frac{d_\Gamma(\boldsymbol{x})}{\varepsilon}, & d_\Gamma(\boldsymbol{x}) \le \varepsilon, \\ 1, & \text{otherwise,} \end{cases}$$

for $\varepsilon > 0$ where d_Γ is a *distance function* given by

$$d_\Gamma(\boldsymbol{x}) := \inf_{\boldsymbol{y} \in \Gamma} |\boldsymbol{x} - \boldsymbol{y}|,$$

with $|\cdot|$ denoting the Euclidean norm in $\mathbb{R}^2$. Note that ψ_ε is zero on Γ and smoothly increases to one off the edge set. Thus $1 - \psi_\varepsilon$ is conversely only supported on a band of

width 2ε around Γ. Hence if Γ is sufficiently smooth we have

$$\int_{\Omega} 1 - \psi_{\varepsilon} d\boldsymbol{x} = \mathcal{O}(\varepsilon) \int_{\Gamma} dS.$$

Note that for smooth sets the length of the curve Γ given by $\int_{\Gamma} dS$ coincides with its one-dimensional Hausdorff measure. Thus ψ_{ε} can indeed be seen as a diffuse representation of the edge set Γ. However, in practice it is of course not possible to give an explicit a priori characterization of a phase function in the manner of (2.7). Instead, one is interested in constructing a functional that is minimized by phase functions which have qualitatively the same properties as the distance function seen in (2.7). Note that the following considerations crucially rely on the fact that ψ_{ε} is a scaled distance function of the form (2.7). We consider the energy associated to ψ_{ε} [She05, Sec. 2]

$$(2.8) \qquad L_{\varepsilon}[\psi_{\varepsilon}] := \int_{\Omega} \varepsilon \left| \nabla \psi_{\varepsilon} \right|^2 + \frac{1}{4\varepsilon} (1 - \psi_{\varepsilon})^2 d\boldsymbol{x}.$$

Let $\eta_{\varepsilon} := (1 - \psi_{\varepsilon})^2$ then Cauchy's inequality (see Appendix B.1) implies

$$(2.9) \qquad L_{\varepsilon}[\psi_{\varepsilon}] \geq \int_{\Omega} \left| \nabla \psi_{\varepsilon} \right| \left| 1 - \psi_{\varepsilon} \right| \, d\boldsymbol{x} = \frac{1}{2} \int_{\Omega} \left| \nabla \eta_{\varepsilon} \right| \, d\boldsymbol{x}.$$

Still assuming sufficient regularity of Γ we may consider it as a smooth curve in Ω. Thus we may parameterize Γ using tangential and normal coordinates t and n respectively. In other words we can write η_{ε} in terms of this new coordinate system as $\eta_{\varepsilon}(t, n)$. By construction η_{ε} remains nearly constant along the tangential direction thus $\left| \frac{\partial}{\partial t} \eta_{\varepsilon} \right| \approx 0$ and hence $\left| \nabla \eta_{\varepsilon} \right| = \sqrt{\frac{\partial}{\partial t} \eta_{\varepsilon}^2 + \frac{\partial}{\partial n} \eta_{\varepsilon}^2} \approx \left| \frac{\partial}{\partial n} \eta_{\varepsilon} \right|$. Recall that η_{ε} is one on Γ and smoothly decreases to zero on a band of width 2ε around Γ. Thus the total variation of η_{ε} the in normal direction is two (η_{ε} increases from zero to one and back to zero when passing from one side of Γ to the other). These considerations motivate the estimate

$$\frac{1}{2} \int_{\Omega} \left| \nabla \eta_{\varepsilon} \right| \, d\boldsymbol{x} \approx \int_{\Gamma} \int_{-\varepsilon}^{\varepsilon} \frac{1}{2} \left| \frac{\partial}{\partial n} \eta_{\varepsilon} \right| \, dn \, dS \approx \int_{\Gamma} dS.$$

This together with (2.9) eventually yields $L_{\varepsilon}[\psi_{\varepsilon}] \geq \int_{\Gamma} dS$. It can be shown [AK06, Sec. 4.2.4] that L_{ε} approaches $\int_{\Gamma} dS$ as $\varepsilon \to 0$. Thus for ε small enough we conclude that

$$L_{\varepsilon}[\psi_{\varepsilon}] \approx \int_{\Gamma} dS.$$

To achieve a similar limiting property without relying on an explicit characterization of ψ_{ε} an additional term is needed. Adding the term $\int_{\Omega} \left| \nabla I \right|^2 \psi d\boldsymbol{x}$ completes the following approximation of the Mumford–Shah functional

$$(2.10) \qquad J_{\mathrm{AT}}[I, \psi] := \frac{\alpha}{2} \int_{\Omega} \left| I - \tilde{I} \right|^2 d\boldsymbol{x} + \frac{\kappa}{2} \int_{\Omega} \left| \nabla I \right|^2 \psi \, d\boldsymbol{x} + L_{\varepsilon}[\psi].$$

This is the so-called *Ambrosio–Tortorelli functional* first introduced in 1990 in the famous paper [AT90]. Note that the second term in J_{AT} induces ψ to be small where $\left| \nabla I \right|$ is large whereas the second term of L_{ε} as given by (2.9) provides an incentive for ψ to be close to one. Thus ψ is induced to be small in the proximity of edges and one anywhere
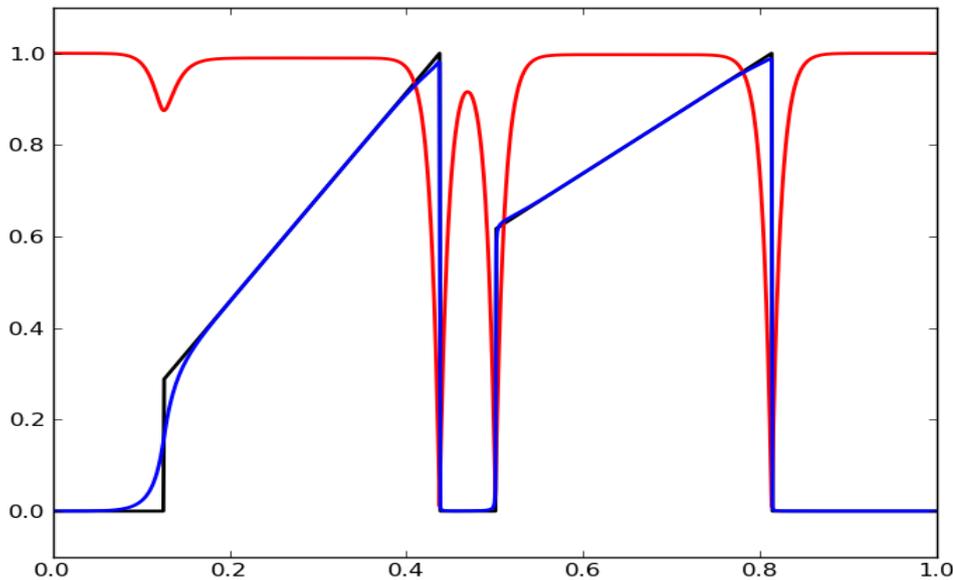
Figure 2.12: The fuzzy nature of Ambrosio–Tortorelli phase functions for a 1D problem. Shown is the Ambrosio–Tortorelli approximation (blue) and phase function (red) for $\tilde{I}$ (black).

else. Hence a phase function minimizing J_{AT} approximates the scaled distance function ψ_ε given by (2.7). Of course, Ambrosio and Tortorelli did not rely on heuristic considerations to illustrate approximation properties of J_{AT}. Using De Giorgi's Γ convergence framework [DG79] they gave a rigorous proof showing convergence of J_{AT} to the Mumford–Shah functional J_{MS} as $\varepsilon \to 0$. Figure 2.11 shows the Ambrosio–Tortorelli segmentation of an exemplary image.

Note, however, that the Ambrosio–Tortorelli approach has some shortcomings as well. Since phase functions minimizing J_{AT} are only smooth approximations to scaled distance functions of the form (2.7), one observes *strong* and *weak* edges. Thus in contrast to K-means an Ambrosio–Tortorelli segmentation does *not* result in a binary edge map but rather in a fuzzy phase function (recall the discussion concerning fuzziness of Chambolle's dual variable in the previous Section).

To illustrate this concept consider the one-dimensional example depicted in Figure 2.12. Here $\tilde{I}$ (in black) is again the ramp function also seen in Figure 2.6. The red curve shows a phase function ψ minimizing J_{AT}. Note that in 1D, Γ is a set of discrete points, the points where $\tilde{I}$ jumps. Thus in this example Γ is a set of four points $\Gamma = \{\gamma_1, \ldots, \gamma_4\}$ ordered from left to right. While the "edges" of $\tilde{I}$ in γ_2, γ_3 and γ_4 are distinctly reflected by considerable drops in the ψ the jump of $\tilde{I}$ in γ_1 is less pronounced. This is a peculiarity of the *discrete* setting; due to the fact that the jump-height of $\tilde{I}$ in γ_1 in relation to the jump magnitudes of $\tilde{I}$ in the rest of Γ is small, ψ manifests decreases of varying magnitudes. Expressed in terms introduced above, $\tilde{I}$ has a weak edge in γ_1. This observation may seem minuscule for the specific example considered as $\psi < 1$ indicates at least the proximity of an edge. However, in the presence of noise it may not be trivial anymore to distinguish between noise and weak edges. Thus in practice one has to rely on heuristic considerations

(a) (b) (c)
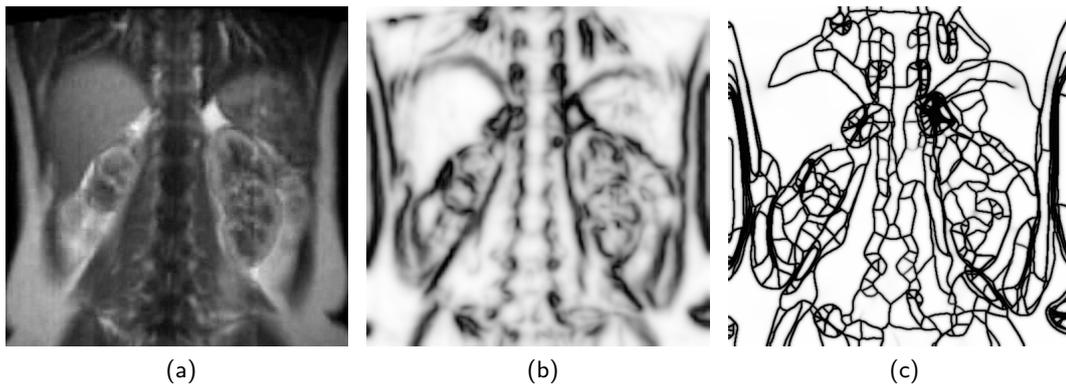
Figure 2.13: An Ambrosio–Tortorelli phase function (b) for the image (a) after thresholding (c).

governing thresholding to convert a phase function ψ to a binary edge map χ. Figure 2.13 illustrates the problems that arise from using these techniques (compare also the discussion in Section 2.2).

A Novel Approach for Computing Binary Edge Maps

Having reviewed the merits and demerits of some segmentation strategies in the previous chapter we postulate the following objectives for our new edge detection approach.

1. Computation of a *binary* edge map χ satisfying

$$\chi = \begin{cases} 0, & \text{on edges,} \\ 1, & \text{otherwise,} \end{cases}$$

   without relying on thresholding techniques.

2. Computation of a smooth approximation of the raw image $\tilde{I}$ albeit *not* imposing piecewise constancy to better account for real data manifesting piecewise smooth intensity variations.

To achieve these goals we combine the strengths of both K-means clustering (Section 2.3) and the Ambrosio–Tortorelli approach (Section 2.5). Note that for the sake of an unobscured presentation of ideas this chapter is solely devoted to modeling and derivation. Rigorous proofs covering existence, uniqueness and well posedness of functionals and minimizers introduced below are given in Chapter 4.

3.1 A Higher Order Approximation of an Image

We start the derivation of our proposed segmentation approach by focusing on Objective 2. Thus in the following we develop a functional that is minimized by a higher order approximation I_s of the raw image $\tilde{I}$. Therefore assume for now that χ is an estimate of a binary edge map satisfying Objective 1. Contrary to the Mumford–Shah approach where $|\Gamma| = 0$, here we will define an edge set in such a way as to have positive measure. Figure 3.1 shows an exemplary χ (in red) for some raw data $\tilde{I}$ (black) in one dimension. Consider the functional

$$\mathcal{F}[I, \chi] := \frac{1}{2} \int_\Omega \left| I - \tilde{I} \right|^2 \chi + \beta \chi \left| \nabla^2 I \right|^2 d\boldsymbol{x},$$

for some $\beta > 0$. By $\nabla^2 I$ we mean the Hessian matrix of I and $\left| \nabla^2 I \right|$ denotes its Frobenius norm (details can be found in Appendix A). The first term of $\mathcal{F}$ is a data fidelity term. However, in contrast to the one seen in, e.g., the Mumford–Shah functional J_MS, it is weighted by the edge map χ. Thus data fidelity is only imposed on $\mathrm{supp}(\chi)$, i.e., only *off edges*. The second term introduces a second order penalty that also acts *off edges*. Thus I is induced to be linear on connected components of $\mathrm{supp}(\chi)$ and hence Objective 2 is
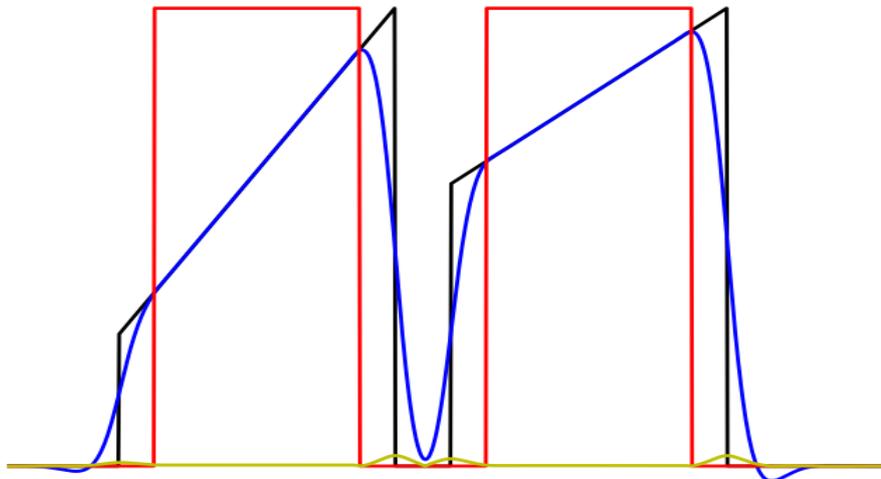
Figure 3.1: Constructing a smooth approximation in 1D. Shown is the data $\tilde{I}$ (black), the estimated edge map χ (red), the smooth approximation I_s (blue) and the fuzzy edge map $\tilde{E}$ (yellow).

accomplished. Note carefully that weighting the penalty by χ prevents smoothing over edges.

Recall, however, that χ is only an estimate of a wanted edge map, i.e., χ might be zero where $\tilde{I}$ exhibits no edge and vice versa (as shown in Figure 3.1). Thus, instead of solely considering the functional $\mathcal{F}$ we define

$$(3.1) \qquad F[I, \chi] := \mathcal{F}[I, \chi] + \delta \mathcal{F}[I, 1],$$

or equivalently

$$F[I, \chi] = \frac{1}{2} \int_\Omega \left| I - \tilde{I} \right|^2 (\delta + \chi) + \beta(\delta + \chi) \left| \nabla^2 I \right|^2 d\boldsymbol{x},$$

for some $0 < \delta \ll 1$. The term $\delta \mathcal{F}[I, 1]$ provides a natural extension of I outside of $\mathrm{supp}(\chi)$. However, since δ is chosen particularly small, F allocates a much higher weight to data fidelity and smoothness on $\mathrm{supp}(\chi)$. Thus for χ fixed I_s defined by

$$I_\mathrm{s} := \arg \min_I F[I, \chi],$$

is a smooth approximation of $\tilde{I}$. Figure 3.1 shows I_s (blue) in a one-dimensional situation. Thus having accomplished Objective 2 we may now focus on Objective 1. Being sufficiently smooth we may determine the wanted edge map by utilizing I_s rather than the possibly noise corrupted $\tilde{I}$. Hence we start by considering the *fuzzy* edge map

$$\tilde{E} := |\nabla I_\mathrm{s}| .$$
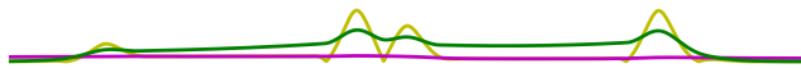
Figure 3.2: Approximating the fuzzy edge map. Shown is the fuzzy edge map $\tilde{E}$ (yellow), the background image E_b (magenta) and the foreground image E_f (green). The plot is scaled for better legibility.

Figure 3.1 shows $\tilde{E}$ in a one-dimensional situation. Note that $\tilde{E}$ is qualitatively similar to an Ambrosio–Tortorelli phase function. However, as mentioned above we do not want to rely on thresholding techniques to extract a binary edge map from $\tilde{E}$. Instead we utilize considerations similar to those that motivated the design of F to compute a *segmentation* of $\tilde{E}$.

3.2 Segmentation of a Fuzzy Edge Map

With a view to $\mathcal{F}$ consider the functional

$$\mathcal{J}[E, \chi] := \frac{1}{2} \int_\Omega \left| E - \tilde{E} \right|^2 \chi + \beta \chi \left| \nabla E \right|^2 d\boldsymbol{x}.$$

Albeit of the same structure note the differences of this functional compared to $\mathcal{F}$. We want to compute a segmentation of the previously obtained fuzzy edge map. Thus the data fidelity term in $\mathcal{J}$ involves $\tilde{E}$ rather than the image $\tilde{I}$. Secondly, in contrast to $\mathcal{F}$, a first order regularity term is used. This follows the idea of a membrane model for $\tilde{E}$; the heuristic concept is to fix membranes from above $\tilde{E}$ and from below $\tilde{E}$ and to use these membranes to define layers according to which $\tilde{E}$ can be binarized. However, we still have to account for the fact that the estimate χ may be imperfect with regards to Objective 1. Hence we define analogously

$$(3.2) \qquad\qquad J[E, \chi] := \mathcal{J}[E, \chi] + \delta \mathcal{J}[E, 1],$$

or equivalently

$$J[E, \chi] = \frac{1}{2} \int_\Omega \left| E - \tilde{E} \right|^2 (\delta + \chi) + \beta (\delta + \chi) \left| \nabla E \right|^2 d\boldsymbol{x}.$$

Then for fixed χ, E_b is given by

$$E_\mathrm{b} := \arg \min_E J[E, \chi],$$

and is an approximation of $\tilde{E}$. Again, weighting by χ yields better approximation quality of E_b *on* supp(χ). Recall that the original goal was to compute an edge map satisfying Objective 1. In order to achieve this we have to design a technique that allows for making a distinct decision whether a point is in an edge set or not. This decision will be made by comparing quantities *inside* and *outside* estimated edge sets. Hence for fixed χ we
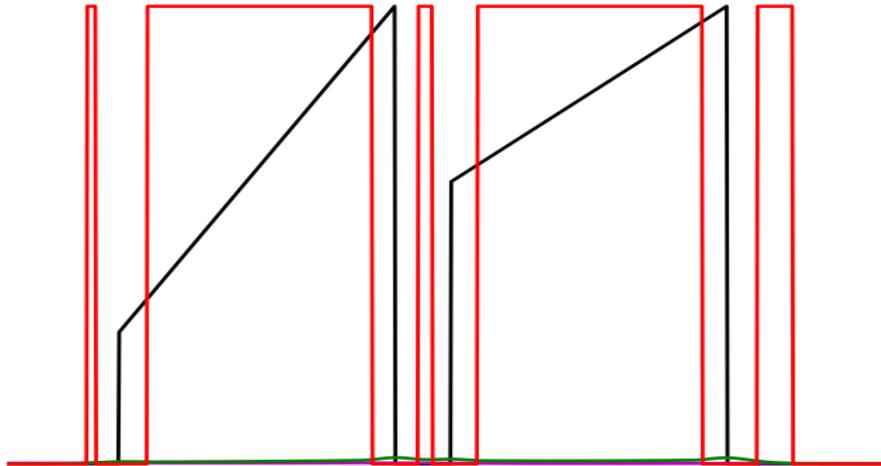
Figure 3.3: An edge map obtained by (3.3). Shown is the raw data $\tilde{I}$ (black), the background image E_b (magenta), the foreground image E_f (green) and the edge map χ (red) as given by (3.3).

introduce

$$E_\mathrm{f} := \arg\min_E J[E, 1 - \chi].$$

Like E_b, this is also an approximation of $\tilde{E}$. In contrast to E_b, however, weighting by $1 - \chi$ induces E_f to estimate $\tilde{E}$ more accurately *off* the support of χ. Thus E_b and E_f are approximations of $\tilde{E}$ outside and inside the estimated edge set respectively. In the following we will refer to E_b as the *background* and to E_f as the *foreground* image. Figure 3.2 clarifies this terminology by referring again to a one-dimensional situation. Shown is the fuzzy edge map $\tilde{E}$ (yellow), the background image E_b (magenta) and the foreground image E_f (green). Note that these curves are obtained by using the estimated edge set and the raw data depicted in Figure 3.1. It can be seen that E_b is below E_f. Hence when looking from top to the bottom E_b is in the "background" whereas E_f lies in the "foreground". Thus from this perspective, E_b approximates $\tilde{E}$ in regions where $\tilde{E}$ is smaller whereas E_f is closer, where $\tilde{E}$ is larger. This observation will be crucial in the following.

3.3 Binarization of a Fuzzy Edge Map

We will exploit the fact the E_b and E_f are two means of approximating $\tilde{E}$. With a view to Figure 3.2 we see that E_f is indeed closer to $\tilde{E}$ in the proximity of jumps whereas E_b approximates $\tilde{E}$ more accurately further away from jumps. This observation motivates the following binarization strategy. Points at which E_b is closer to $\tilde{E}$ are not points of an edge sets. Conversely, points at which E_f is closer to $\tilde{E}$ lie in an edge set. Thus we define
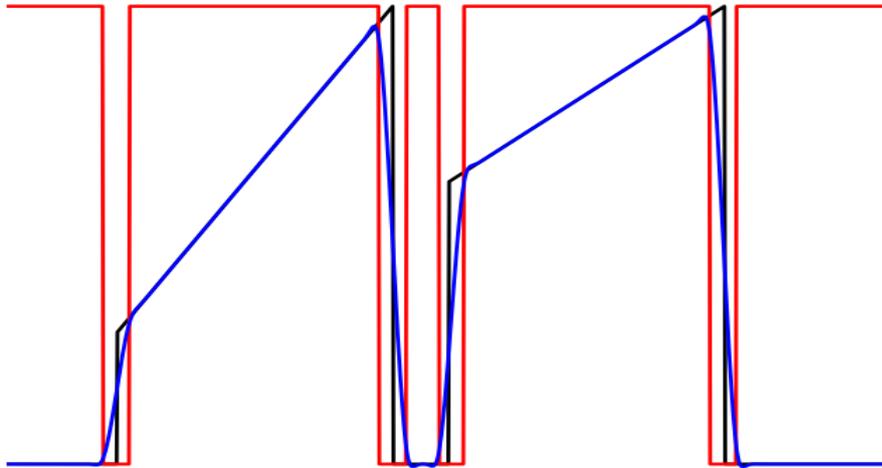
Figure 3.4: Result of Algorithm 3.1 for a one-dimensional problem. Shown is the raw data $\tilde{I}$ (black), the smooth image I_s (blue) and the binary edge map χ (red).

an edge map by

$$(3.3) \qquad \chi(\boldsymbol{x}) := \begin{cases} 0, & \vartheta \left| E_\mathrm{b}(\boldsymbol{x}) - \tilde{E}(\boldsymbol{x}) \right| \leq \left| E_\mathrm{f}(\boldsymbol{x}) - \tilde{E}(\boldsymbol{x}) \right|, \\ 1, & \text{otherwise,} \end{cases}$$

where the parameter $\vartheta \geq 0$ governs the *thickness* of edges. Hence for $\vartheta = 0$ we would obtain $\chi = 0$ everywhere on Ω, i.e., no edges, whereas larger values of ϑ yield thicker edges. A numerical investigation illustrating the effect of ϑ is given in Section 6.4. Figure 3.3 shows an edge map obtained by (3.3). Note that the initial estimated edge map used to compute I_s and $\tilde{E}$, and thus E_b and E_f, has of course an influence on the form of χ given by (3.3). Thus to eliminate the possible negative influence of a bad initial edge map we refine χ iteratively. This idea is the foundation of the algorithmic approach detailed in the next section.

3.4 An Algorithm to Compute Binary Edge Maps

Note that all considerations presented so far crucially rely on the fact that some estimate for a binary edge map satisfying Objective 1 is available. Thus, in practice the first question that arises is how to obtain such an estimate. Extensive numerical tests in one as well as in two dimensions indicated that edge maps obtained by K-means clustering with $K = 2$ are the most robust initial guesses. Thus for χ being a K-means edge map, first the smooth image I_s and then the fuzzy edge map $\tilde{E}$ are computed. Subsequently, $\tilde{E}$ is used to compute the back- and foreground images E_b and E_f respectively. In the next step χ is updated according to (3.3). The crucial idea now is not to go back to recompute I_s

---

Algorithm 3.1 Computation of binary edge maps

Input: $\tilde{I}$, β, δ, ϑ
Output: χ, I_s

1: Compute an initial edge map χ using K-means clustering (see Section 2.3) with $K = 2$

2: **while** χ changes **do**
3:      Compute $I_\text{s} = \arg\min_I F[I, \chi]$
4:      Set $\tilde{E} = |\nabla I_\text{s}|$
5:      **while** χ changes **do**
6:          Compute $E_b = \arg\min_E J[E, \chi]$ and $E_\text{f} = \arg\min_E J[E, 1 - \chi]$
7:          Set

$$\chi(\boldsymbol{x}) := \begin{cases} 0, & \vartheta \left| E_\text{b}(\boldsymbol{x}) - \tilde{E}(\boldsymbol{x}) \right| \leq \left| E_\text{f}(\boldsymbol{x}) - \tilde{E}(\boldsymbol{x}) \right|, \\ 1, & \text{otherwise.} \end{cases}$$

8:      **end while**
9: **end while**
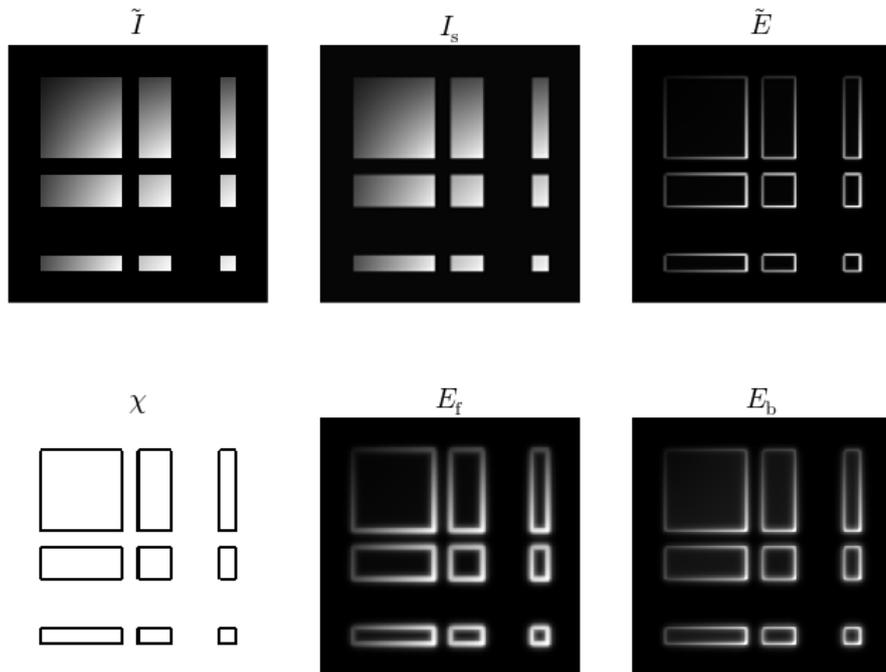
Figure 3.5: Result of Algorithm 3.1 for an artificial image. Shown is the raw image $\tilde{I}$, the smooth approximation I_s, the fuzzy edge map $\tilde{E}$, the binary edge map χ, the foreground image E_f and the background image E_b.

right away but rather update E_{b} and E_{f} using the new edge map χ. This is repeated until either changes in E_{f} are sufficiently small or χ does not change any more. We will refer to this procedure as the *Inner Iteration*. Only after the Inner Iteration has converged are I_{s} and hence $\tilde{E}$ updated. Then using the recomputed fuzzy edge map the Inner Iteration is started again. We call this the *Outer Iteration*. Algorithm 3.1 summarizes the procedure. Note that Algorithm 3.1 just sketches the basic mechanics of the developed approach. A more detailed pseudo-code addressing stopping criteria and specific discretization and implementation details is given in Section 5.3.

Figure 3.4 depicts the result of this algorithmic approach for the one-dimensional model problem used throughout this chapter. Figure 3.5 shows the result of Algorithm 3.1 for a two-dimensional artificial image.

As stated in the beginning of this chapter, many mathematical details have been omitted deliberately here to allow the reader to focus on the considerations that govern the design of Algorithm 3.1. The next chapter subsequently fills those gaps and presents a rigorous analysis of the presented approach.

Analysis of the Proposed Method

This chapter is devoted to an extensive analysis of the algorithmic strategy introduced in Chapter 3. Note that for reasons of technical necesisty some mathematical details deviate from the formulation seen in Algorithm 3.1. However, these adaptations have been introduced with considerable precautions. For the sake of brevity and to establish a consistent basis for the theoretical analysis below, assume the following.

Assumption 1. *Let $\Omega = (0,1)^2$ and $\tilde{I} \in L^2(\Omega)$. Further, let $0 < \delta \ll 1$, $\beta > 0$ and $\vartheta > 0$.*

Note further that in the earlier discussion we relied on an intuitive notion of an edge map. In a more rigorous mathematical framework we have to specify what we mean by an edge map. Thus in the following we will rather consider characteristic functions of "appropriate" sets. More specifically we only consider characteristic functions of sets in the metric space $\mathcal{M}(\Omega)$ (defined in Appendix B.2). In other words, here, the term *characteristic function* implies measurability since $\mathcal{M}(\Omega)$ solely consists of Lebesgue measurable sets. In the course of the following sections the reason to start this analysis with the Inner Iteration will become clear.

4.1 The Inner Iteration

Given Assumption 1 let χ be a characteristic function of some subset of Ω and let $E \in H^1(\Omega)$. Recall the functional J given by (3.2). We want to deduce the necessary optimality conditions for the minimization of J with respect to E for a fixed χ. To simplify notation we define

$$(4.1) \qquad J_{\mathrm{b}}[E] := J[E, \chi],$$

and thus consider the minimization problem

$$(4.2) \qquad \inf_{E \in H^1(\Omega)} J_{\mathrm{b}}[E].$$

Note that J_{b} is everywhere Gâteaux differentiable. Thus we compute the Gâteaux derivative (see Appendix B.2) of J_{b} in an arbitrary direction $v \in C^\infty(\bar{\Omega})$

$$\frac{\delta J_{\mathrm{b}}}{\delta E}[E; v] = \frac{d}{ds}\left(\frac{1}{2}\int_\Omega \left|E + sv - \tilde{E}\right|^2 (\delta + \chi) + \beta(\delta + \chi)\left|\nabla(E + sv)\right|^2 dx\right)\Bigg|_{s=0}$$
$$= \int_\Omega (E + sv - \tilde{E})(\delta + \chi)v + \beta(\delta + \chi)\nabla(E + sv) \cdot \nabla v \, dx\Bigg|_{s=0}$$
$$= \int_\Omega (E - \tilde{E})(\delta + \chi)v + \beta(\delta + \chi)\nabla E \cdot \nabla v \, dx.$$

Hence the weak formulation of the necessary optimality conditions for (4.2) is given by (compare, e.g., [Lue69, Sec. 7.4, Th. 1])

$$(4.3) \qquad\qquad B_{\mathrm{b}}[u, v] = f_{\mathrm{b}}[v], \quad \forall v \in H^1(\Omega),$$

where $B_{\mathrm{b}} : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ is a bilinear form defined by

$$(4.4) \qquad\qquad B_{\mathrm{b}}[u, v] := \int_\Omega \beta(\delta + \chi)\nabla u \cdot \nabla v + (\delta + \chi)uv \, dx,$$

and $f_{\mathrm{b}} : H^1(\Omega) \to \mathbb{R}$ is a linear functional

$$(4.5) \qquad\qquad f_{\mathrm{b}}[v] := \int_\Omega \tilde{E}(\delta + \chi)v \, dx.$$

Similarly we set

$$(4.6) \qquad\qquad J_{\mathrm{f}}[E] := J[E, 1 - \chi],$$

and thus obtain the minimization problem

$$(4.7) \qquad\qquad \inf_{E \in H^1(\Omega)} J_{\mathrm{f}}[E].$$

We compute again

$$\frac{\delta J_{\mathrm{f}}}{\delta E}[E; v] = \int_\Omega (E - \tilde{E})(\delta + 1 - \chi)v + \beta(\delta + 1 - \chi)\nabla E \cdot \nabla v \, dx.$$

Thus the weak formulation of the necessary optimality conditions for (4.7) is analogously given by

$$(4.8) \qquad\qquad B_{\mathrm{f}}[u, v] = f_{\mathrm{f}}[v], \quad \forall v \in H^1(\Omega),$$

with the bilinear form $B_{\mathrm{f}} : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ defined through

$$(4.9) \qquad\qquad B_{\mathrm{f}}[u, v] := \int_\Omega \beta(\delta + 1 - \chi)\nabla u \cdot \nabla v + (\delta + 1 - \chi)uv \, dx,$$

and the linear functional $f_{\mathrm{f}} : H^1(\Omega) \to \mathbb{R}$

$$(4.10) \qquad\qquad f_{\mathrm{f}}[v] := \int_\Omega \tilde{E}(\delta + 1 - \chi)v \, dx.$$

First we show that the minimization problems (4.2) and (4.7) both have unique solutions.

Theorem 1. *Given Assumption 1 let $\tilde{E} \in L^2(\Omega)$ and χ be a characteristic function on Ω. Then the minimization problems (4.2) and (4.7) respectively have unique solutions.*

Proof. We want to apply the Lax–Milgram Lemma (Appendix B.2). Both f_b and f_f are bounded since (note that $\|\chi\|_{L^\infty(\Omega)}, \|1-\chi\|_{L^\infty(\Omega)} \leq 1$)

$$|f_\mathrm{b}[v]| \leq \|\delta + \chi\|_{L^\infty(\Omega)} \int_\Omega \left|\tilde{E}v\right| d\boldsymbol{x} \leq (\delta + 1) \left\|\tilde{E}\right\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)},$$

and

$$|f_\mathrm{f}[v]| \leq \|\delta + 1 - \chi\|_{L^\infty(\Omega)} \int_\Omega \left|\tilde{E}v\right| d\boldsymbol{x} \leq (\delta + 1) \left\|\tilde{E}\right\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)}.$$

Next we show that the bilinear forms B_b and B_f are bounded. Starting with B_b we compute

$$\begin{aligned}
|B_\mathrm{b}[u,v]| &\leq \|\beta(\delta + \chi)\|_{L^\infty(\Omega)} \int_\Omega |\nabla u| \, |\nabla v| \, d\boldsymbol{x} + \|\delta + \chi\|_{L^\infty(\Omega)} \int_\Omega |u||v| d\boldsymbol{x} \\
&\leq \beta(\delta + \|\chi\|_{L^\infty(\Omega)}) \|\nabla u\|_{L^2(\Omega)} \|\nabla u\|_{L^2(\Omega)} + (\delta + \|\chi\|_{L^\infty(\Omega)}) \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\
&\leq (\delta + 1)(\beta + 1) \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)},
\end{aligned}$$

where we used the Cauchy–Schwarz inequality (Appendix B.1). Analogously we obtain

$$\begin{aligned}
|B_\mathrm{f}[u,v]| &\leq \beta(\delta + \|1 - \chi\|_{L^\infty(\Omega)}) \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \\
&\quad + (\delta + \|1 - \chi\|_{L^\infty(\Omega)}) \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \\
&= (\delta + 1)(\beta + 1) \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}.
\end{aligned}$$

Left to show is coercivity of both B_b and B_f. Starting again with B_b we get since $\chi \geq 0$ and $\beta, \delta \geq 0$

$$
\begin{aligned}
B_\mathrm{b}[u,u] &= \int_\Omega \beta\,(\delta + \chi)\,\nabla u \cdot \nabla u + (\delta + \chi)u^2 d\boldsymbol{x} \\
&\geq \delta \int_\Omega \beta \nabla u^2 + u^2 d\boldsymbol{x} \\
&\geq \delta \min\{\beta, 1\} \|u\|_{H^1(\Omega)}^2,
\end{aligned}
$$

(4.11)

and similarly

$$B_\mathrm{f}[u,u] \geq \delta \min\{\beta, 1\} \|u\|_{H^1(\Omega)}^2.$$

Thus both B_b and B_f are coercive on $H^1(\Omega)$ and the conditions of the Lax–Milgram Lemma are satisfied which implies that each weak formulation (4.3) and (4.8) has a unique solution. This together with the observation that both cost functionals J_b and J_f are strictly convex in E yields that the minimization problems (4.2) and (4.7) respectively have unique solutions [ET99, Chap. 2 Proposition 1.2]. □

We are primarily interested in the effect that changes in the edge map χ have on the back- and foreground images and vice versa. Thus in the following we will investigate continuity of these dependencies. Therefore it is crucial to choose a notion of convergence of characteristic functions that is tightly connected to convergence of the associated sets. It turns

out that the strong $L^1(\Omega)$-topology is a natural choice when working with characteristic functions. Conversely a very convenient way to define convergence of a sequence of sets $\{\Omega_k\}_{k\geq 1}$ to some limit set Ω_χ is to use the measure of the symmetric difference of sets $|\Omega_\chi \vartriangle \Omega_k|$ (for details see Appendix B.2). The following Lemma shows that these two modes of convergence are equivalent (a similar result is presented in [CFK04]).

Lemma 1. *Given Assumption 1 let $\{\chi_k\}_{k\geq 1}$ be a sequence of characteristic functions on Ω and $\Omega_k := \mathrm{supp}(\chi_k)$ for all $k \in \mathbb{N}$. Similarly let χ be also a characteristic function on Ω with $\Omega_\chi := \mathrm{supp}(\chi)$. Then*

$$|\Omega_\chi \vartriangle \Omega_k| \overset{k\to\infty}{\longrightarrow} 0 \quad \Leftrightarrow \quad \|\chi - \chi_k\|_{L^1(\Omega)} \overset{k\to\infty}{\longrightarrow} 0.$$

Proof. Every χ_k and χ are characteristic which implies

$$\Omega_\chi \backslash \Omega_k = \{\boldsymbol{x} \in \Omega \,|\, \chi(\boldsymbol{x}) = 1 \wedge \chi_k(\boldsymbol{x}) = 0\},$$

and analogously

$$\Omega_k \backslash \Omega_\chi = \{\boldsymbol{x} \in \Omega \,|\, \chi(\boldsymbol{x}) = 0 \wedge \chi_k(\boldsymbol{x}) = 1\}.$$

Since $\Omega_\chi \backslash \Omega_k$ and $\Omega_k \backslash \Omega_\chi$ are disjoint we obtain

$$
\begin{aligned}
|\Omega_\chi \vartriangle \Omega_k| &= |(\Omega_\chi \backslash \Omega_k) \cup (\Omega_k \backslash \Omega_\chi)| \\
&= |\{\boldsymbol{x} \in \Omega \,|\, \chi(\boldsymbol{x}) = 1 \wedge \chi_k(\boldsymbol{x}) = 0\}| + |\{\boldsymbol{x} \in \Omega \,|\, \chi(\boldsymbol{x}) = 0 \wedge \chi_k(\boldsymbol{x}) = 1\}| \\
&= \int_\Omega \chi(1 - \chi_k)\, d\boldsymbol{x} + \int_\Omega \chi_k(1 - \chi)\, d\boldsymbol{x} \\
&= \int_\Omega \chi - 2\chi\chi_k + \chi_k\, d\boldsymbol{x} = \int_\Omega \chi^2 - 2\chi\chi_k + \chi_k^2\, d\boldsymbol{x} = \int_\Omega |\chi - \chi_k|^2\, d\boldsymbol{x} \\
&= \int_\Omega |\chi - \chi_k|\, d\boldsymbol{x} \\
&= \|\chi - \chi_k\|_{L^1(\Omega)},
\end{aligned}
$$

which proves the claim. $\qquad\square$

Let

(4.12) $$E_{\mathrm{b}} := \arg \min_{E \in H^1(\Omega)} J_{\mathrm{b}}[E] \text{ and } E_{\mathrm{f}} := \arg \min_{E \in H^1(\Omega)} J_{\mathrm{f}}[E],$$

with the functionals given by (4.1) and (4.6) respectively. Note that Theorem 1 guarantees well posedness of this definition. Now that we have selected and characterized a suitable notion of convergence for characteristic functions (and the associated sets) we can investigate the mutual influence of fore- and background images on edge maps and vice versa. We need the following technical result first.

Lemma 2. *Given Assumption 1 define*

(4.13)
$$\varphi_\delta(\boldsymbol{x}) := \frac{1}{4\pi\delta} e^{-\frac{|\boldsymbol{x}|^2}{4\delta}},$$

and for any $g \in L^p(\Omega)$ with $1 \le p < \infty$ let

$$g_\delta := g * \varphi_\delta = \int_{\mathbb{R}^2} \varphi_\delta(\boldsymbol{x} - \boldsymbol{y}) g(\boldsymbol{y}) \, d\boldsymbol{y},$$

where g is extended by 0 outside of Ω. Then the following statements hold.

1. *The function g_δ is real analytic on Ω and $g_\delta \xrightarrow{a.e.} g$. Moreover, if $0 \le g \le 1$ then $0 \le g_\delta \le 1$.*

2. *Let $\omega \subset \Omega$ with $|\omega| > 0$ and assume $g_\delta \equiv \gamma$ on ω for some $\gamma \in \mathbb{R}$. Then $\varphi_\delta \equiv \gamma$ on Ω.*

3. *For $1 \le p < \infty$ choose $1 \le r \le \infty$ such that $1/r + 1 - 1/p \in [0, 1]$. Then the operator*

(4.14)
$$R_\delta : \begin{cases} L^p(\Omega) & \to L^r(\Omega), \\ g & \mapsto g * \varphi_\delta, \end{cases}$$

is continuous and injective.

Proof. We prove each claim separately

1. Observe that $\varphi_\delta(\boldsymbol{x} - \boldsymbol{y})$ is the two-dimensional heat kernel and thus φ_δ as well as g_δ are real analytic functions on $\mathbb{R}^2$ and hence also on Ω [Byu98]. Thus $G(\delta, \boldsymbol{x}) := g_\delta(\boldsymbol{x})$ solves

$$\begin{cases} \dfrac{\partial}{\partial \delta} G + \Delta G = 0, & \boldsymbol{x} \in \mathbb{R}^2, \delta > 0, \\ G(0, \boldsymbol{x}) = g(\boldsymbol{x}), & \boldsymbol{x} \in \mathbb{R}^2, \end{cases}$$

   and hence $\lim_{\delta \to 0} G(\boldsymbol{x}, \delta) = g(\boldsymbol{x})$ for almost every $\boldsymbol{x}$ (compare, e.g., [AST12]). Suppose now $0 \le g \le 1$. Then since $\varphi_\delta \ge 0$ it follows that $g_\delta \ge 0$ and further

$$|g_\delta(\boldsymbol{x})| = \left| \int_{\mathbb{R}^2} \varphi_\delta(\boldsymbol{y}) g(\boldsymbol{x} - \boldsymbol{y}) \, d\boldsymbol{y} \right| \le 1 \cdot \int_{\mathbb{R}^2} \varphi_\delta(\boldsymbol{y}) \, d\boldsymbol{y} = 1,$$

   since φ_δ has integral one on $\mathbb{R}^2$.

2. The second claim follows from real analyticity of g_δ on Ω (compare [Kra01, Cor.2.38]).

3. Since φ_δ is real analytic on $\mathbb{R}^2$ it is in $L^q(\Omega)$ for any $1 \le q \le \infty$. Thus for $g \in L^p(\Omega)$ fix $1 \le r \le \infty$ such that $1/r + 1 - 1/p = 1/q$. Then by Young's inequality for convolutions (Appendix B.1) we have

$$\|R_\delta[g]\|_{L^r(\Omega)} \le \|g\|_{L^p(\Omega)} \|\varphi_\delta\|_{L^q(\Omega)},$$

   thus R_δ is well defined and continuous. Since $\int_{\mathbb{R}^2} \varphi_\delta d\boldsymbol{x} = 1$ the operator R_δ is also injective (see, e.g., [GLV09]).

$\square$

We show in the following that sets which are closely related to the edge maps seen in Algorithm 3.1 (compare Remark 1) are converging provided that the underlying images converge.

Lemma 3. *Given Assumption 1 and $\tilde{E} \in H^1(\Omega)$ let $\{E_{b_k}\}_{k \geq 1}$ and $\{E_{f_k}\}_{k \geq 1}$ be two sequences strongly converging in $H^1(\Omega)$ to some $E_b^\star$ and $E_f^\star$ respectively. Further, let $Z \in L^2(\Omega) \backslash H^1(\Omega)$ and φ_δ be given by (4.13). Define*

$$\Omega_k := \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\sqrt{\left| E_{f_k} - \tilde{E} \right|^2 + \delta} - \vartheta \sqrt{\left| E_{b_k} - \tilde{E} \right|^2 + \delta} + \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) \geq 0 \right. \right\},$$

and

$$\Omega_\chi := \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\sqrt{\left| E_f^\star - \tilde{E} \right|^2 + \delta} - \vartheta \sqrt{\left| E_b^\star - \tilde{E} \right|^2 + \delta} + \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) \geq 0 \right. \right\}.$$

Then $|\Omega_\chi \triangle \Omega_k| \stackrel{k \to \infty}{\longrightarrow} 0$.

Proof. We introduce the following notation

$$h_k := \vartheta \sqrt{\left| E_{b_k} - \tilde{E} \right|^2 + \delta}, \qquad h := \vartheta \sqrt{\left| E_b^\star - \tilde{E} \right|^2 + \delta},$$

$$g_k := \sqrt{\left| E_{f_k} - \tilde{E} \right|^2 + \delta}, \qquad g := \sqrt{\left| E_f^\star - \tilde{E} \right|^2 + \delta}.$$

Observe that

$$\int_\Omega g^2 \, d\boldsymbol{x} = \int_\Omega \left| E_f^\star - \tilde{E} \right|^2 + \delta \, d\boldsymbol{x} = \left\| E_f^\star - \tilde{E} \right\|_{L^2(\Omega)}^2 + \delta \, |\Omega| < \infty,$$

since Ω is bounded and $E_f^\star$ just as $\tilde{E}$ are in $H^1(\Omega) \subset L^2(\Omega)$ by assumption. Thus $g \in L^2(\Omega)$. Similarly we obtain

$$\|\nabla g\|_{L^2(\Omega)} = \left\| \frac{(E_f - \tilde{E})(\nabla E_f - \nabla \tilde{E})}{\sqrt{\left| E_f - \tilde{E} \right|^2 + \delta}} \right\|_{L^2(\Omega)} \leq \frac{1}{\sqrt{\delta}} \left\| (E_f - \tilde{E})(\nabla E_f - \nabla \tilde{E}) \right\|_{L^2(\Omega)} < \infty,$$

and thus $g \in H^1(\Omega)$. Analogously one may show that g_k, h_k and h are also in $H^1(\Omega)$. Now we rewrite the definitions of Ω_k and Ω_χ as follows

$$(4.15) \qquad \begin{aligned} \Omega_k &= \left\{ \boldsymbol{x} \in \Omega \left| ((g_k - h_k + \delta Z) * \varphi_\delta)(\boldsymbol{x}) \geq 0 \right. \right\}, \\ \Omega_\chi &= \left\{ \boldsymbol{x} \in \Omega \left| ((g - h + \delta Z) * \varphi_\delta)(\boldsymbol{x}) \geq 0 \right. \right\}. \end{aligned}$$

Thus the set $\Omega_\chi \backslash \Omega_k$ is given by

$$(\Omega_\chi \backslash \Omega_k) = \left\{ \boldsymbol{x} \in \Omega \left| ((g - h + \delta Z) * \varphi_\delta)(\boldsymbol{x}) \geq 0 \wedge ((h_k - g_k - \delta Z) * \varphi_\delta)(\boldsymbol{x}) > 0 \right. \right\}.$$

Let $\varepsilon > 0$ and define

$$U_k(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega \,|\, \left((g - h + h_k - g_k) * \varphi_\delta \right)(\boldsymbol{x}) \geq \varepsilon \right\},$$

then by distributivity of the convolution it follows immediately that $(\Omega_\chi \backslash \Omega_k) \subseteq U_k(0)$. Introducing the sets

$$(4.16) \qquad \begin{aligned} V(\varepsilon) := \{ \boldsymbol{x} \in \Omega \,|\, & 0 \leq \left((g - h + \delta Z) * \varphi_\delta \right)(\boldsymbol{x}) < \varepsilon \\ & \wedge \left((h_k - g_k - \delta Z) * \varphi_\delta \right)(\boldsymbol{x}) > 0 \}, \end{aligned}$$

and

$$W(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega \,|\, \left((g - h + \delta Z) * \varphi_\delta \right)(\boldsymbol{x}) \geq \varepsilon \wedge \left((h_k - g_k - \delta Z) * \varphi_\delta \right)(\boldsymbol{x}) > 0 \right\},$$

we can rewrite $\Omega_\chi \backslash \Omega_k$ so that

$$(4.17) \qquad (\Omega_\chi \backslash \Omega_k) = (V(\varepsilon) \cup W(\varepsilon)).$$

Obviously

$$(4.18) \qquad |W(\varepsilon)| \leq |U_k(\varepsilon)|, \quad \forall \varepsilon > 0,$$

and we show in the following that $|V(\varepsilon)| \to 0$ as $\varepsilon \to 0$. Thus suppose $(\Omega_\chi \backslash \Omega_k) \neq \emptyset$ (if $(\Omega_\chi \backslash \Omega_k) = \emptyset$ then by (4.17), $V(\varepsilon) = \emptyset$ for all $\varepsilon \geq 0$ and the claim is trivial). If $\{\varepsilon^k\}_{k \geq 1}$ is a monotonically decreasing null sequence then the Definition (4.16) of $V(\varepsilon)$ implies that $V(\varepsilon^{k+1}) \subset V(\varepsilon^k)$ since $\varepsilon^{k+1} < \varepsilon^k$. Hence $\{V(\varepsilon^k)\}_{k>0}$ is a decreasing nested sequence of sets such that $V_0 := \bigcap_{k=1}^{\infty} V(\varepsilon^k)$ and $|V(\varepsilon)| \overset{\varepsilon \to 0}{\longrightarrow} |V(0)|$. We want to show that $|V_0| = 0$. Assume for the sake of contradiction that V_0 has positive measure and hence

$$(4.19) \qquad (g - h + \delta Z) * \varphi_\delta = 0 \text{ on } V_0.$$

According to Lemma 2 $(g - h + \delta Z) * \varphi_\delta$ is real analytic on Ω. Since $|V_0| > 0$ relation (4.19) implies again by Lemma 2 that

$$(4.20) \qquad (g - h + \delta Z) * \varphi_\delta = 0 \text{ on } \Omega.$$

Let R_δ be the operator defined in (4.14) then (4.20) can be rewritten as

$$R_\delta[g - h + \delta Z] = 0 \text{ on } \Omega.$$

Since $R_\delta[0] = 0$ as well we have

$$0 = R_\delta[g - h + \delta Z] = R_\delta[0].$$

Lemma 2 also guarantees that R_δ is injective which thus implies

$$g - h + \delta Z = 0 \text{ on } \Omega,$$

or equivalently

$$h - g = \delta Z \text{ on } \Omega.$$

However, this is a contradiction since by assumption $Z \in L^2(\Omega) \backslash H^1(\Omega)$ but $h - g \in H^1(\Omega)$. Thus V_0 cannot have positive measure and hence $|V(\varepsilon)| \overset{\varepsilon \to 0}{\longrightarrow} |V(0)| = 0$. This means

$$(4.21) \qquad \forall \varepsilon > 0 \, \exists \varepsilon_1 > 0 : V(\hat{\varepsilon}) < \frac{\varepsilon}{4}, \quad \forall \hat{\varepsilon} \in (0, \varepsilon_1).$$

Further, by employing Markov's inequality (Appendix B.1) and again distributivity of the convolution we obtain the following estimate

$$(4.22) \qquad \begin{aligned} |U_k(\varepsilon)| \leq & \frac{1}{\varepsilon} \int_\Omega |(g - h + h_k - g_k) * \varphi_\delta| \, d\boldsymbol{x} \\ \leq & \frac{1}{\varepsilon} \left(\|(g - g_k) * \varphi_\delta\|_{L^1(\Omega)} + \|(h_k - h) * \varphi_\delta\|_{L^1(\Omega)} \right), \quad \forall \varepsilon > 0. \end{aligned}$$

Note that for some positive real numbers α and α_k we have the identity

$$\alpha - \alpha_k = (\sqrt{\alpha + \delta} + \sqrt{\alpha_k + \delta})(\sqrt{\alpha + \delta} - \sqrt{\alpha_k + \delta}),$$

or equivalently

$$(4.23) \qquad \sqrt{\alpha + \delta} - \sqrt{\alpha_k + \delta} = \frac{\alpha - \alpha_k}{\sqrt{\alpha + \delta} + \sqrt{\alpha_k + \delta}}.$$

Now let $\alpha = \left| E_{\mathrm{b}}^\star - \tilde{E} \right|^2$ and $\alpha_k := \left| E_{\mathrm{b}_k} - \tilde{E} \right|^2$ then we compute

$$(4.24) \qquad \begin{aligned} \|h - h_k\|_{L^1(\Omega)} &= \int_\Omega \vartheta \left| \sqrt{\left| E_{\mathrm{b}}^\star - \tilde{E} \right|^2 + \delta} - \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E} \right|^2 + \delta} \right| d\boldsymbol{x} \\ &= \int_\Omega \vartheta \left| \frac{\left| E_{\mathrm{b}}^\star - \tilde{E} \right|^2 - \left| E_{\mathrm{b}_k} - \tilde{E} \right|^2}{\sqrt{\left| E_{\mathrm{b}}^\star - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E} \right|^2 + \delta}} \right| d\boldsymbol{x} \\ &= \int_\Omega \vartheta \left| \frac{(E_{\mathrm{b}_k} - E_{\mathrm{b}}^\star)(E_{\mathrm{b}_k} + E_{\mathrm{b}}^\star - 2\tilde{E})}{\sqrt{\left| E_{\mathrm{b}}^\star - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E} \right|^2 + \delta}} \right| d\boldsymbol{x} \\ &\leq \vartheta \, \|E_{\mathrm{b}_k} - E_{\mathrm{b}}^\star\|_{L^2(\Omega)} \left\| \frac{E_{\mathrm{b}_k} + E_{\mathrm{b}}^\star - 2\tilde{E}}{\sqrt{\left| E_{\mathrm{b}}^\star - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E} \right|^2 + \delta}} \right\|_{L^2(\Omega)}. \end{aligned}$$

By assumption $E_{\mathrm{b}_k} \underset{H^1(\Omega)}{\longrightarrow} E_{\mathrm{b}}^\star$ and hence $E_{\mathrm{b}_k} \underset{L^2(\Omega)}{\longrightarrow} E_{\mathrm{b}}^\star$. Thus for every $\varepsilon > 0$ we can pick $K \in \mathbb{N}$ such that for $k \geq K$ we have $\|E_{\mathrm{b}_k}\|_{L^2(\Omega)} < \|E_{\mathrm{b}}^\star\|_{L^2(\Omega)} + \varepsilon$. Hence we obtain

$$(4.25) \qquad \left\| \frac{E_{\mathrm{b}_k} + E_{\mathrm{b}}^\star - 2\tilde{E}}{\sqrt{\left| E_{\mathrm{b}}^\star - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E} \right|^2 + \delta}} \right\|_{L^2(\Omega)} \leq \frac{1}{\delta} \left(\|E_{\mathrm{b}}^\star\|_{L^2(\Omega)} + \left\| \tilde{E} \right\|_{L^2(\Omega)} + \frac{\varepsilon}{2} \right).$$

Thus by combining (4.24) and (4.25) and using Young's inequality for convolutions (Appendix B.1) we get

$$\|(h - h_k) * \varphi_\delta\|_{L^1(\Omega)} \le \|h - h_k\|_{L^1(\Omega)} \|\varphi_\delta\|_{L^1(\Omega)}$$

(4.26)
$$\le \frac{\vartheta}{\delta} \left\| E_{b_k} - \tilde{E} \right\|_{L^2(\Omega)} \times$$
$$\left(\|E_b^\star\|_{L^2(\Omega)} + \left\| \tilde{E} \right\|_{L^2(\Omega)} + \frac{\varepsilon}{2} \right) \|\varphi_\delta\|_{L^1(\Omega)} \xrightarrow{k \to \infty} 0.$$

Similar considerations show that also

(4.27)
$$\|(g - g_k) * \varphi_\delta\|_{L^1(\Omega)} \xrightarrow{k \to \infty} 0,$$

since $\{E_{f_k}\}_{k \ge 1}$ converges strongly in $H^1(\Omega)$ as well. Thus due to (4.22) we get

(4.28)
$$|U_k(\varepsilon)| \xrightarrow{k \to \infty} 0, \quad \forall \varepsilon > 0.$$

Using ε_1 appearing in (4.21) we infer from (4.28) the existence of $K_1 \in \mathbb{N}$ such that

(4.29)
$$|U_k(\hat{\varepsilon})| < \frac{\varepsilon}{4}, \quad \forall k \ge K_1, \forall \hat{\varepsilon} \in (0, \varepsilon_1).$$

Thus combining (4.17), (4.18), (4.21) and (4.29) yields

(4.30)
$$|\Omega_\chi \backslash \Omega_k| = |W(\hat{\varepsilon}) \cup V(\hat{\varepsilon})| \le |U_k(\hat{\varepsilon})| + |V(\hat{\varepsilon})| < \frac{\varepsilon}{2}, \quad \forall k \ge K_1, \forall \hat{\varepsilon} \in (0, \varepsilon_1).$$

Conversely for

$$(\Omega_k \backslash \Omega_\chi) = \{\boldsymbol{x} \in \Omega \,|\, ((h - g - \delta Z) * \varphi_\delta)(\boldsymbol{x}) > 0 \wedge ((g_k - h_k + \delta Z) * \varphi_\delta)(\boldsymbol{x}) \ge 0\}.$$

we define

$$\bar{U}_k(\varepsilon) := \{\boldsymbol{x} \in \Omega \,|\, ((h - g + g_k - h_k) * \varphi_\delta)(\boldsymbol{x}) \ge \varepsilon\},$$

for $\varepsilon > 0$. We get similarly $(\Omega_k \backslash \Omega_\chi) \subseteq \bar{U}_k(0)$ and introduce

$$\bar{V}(\varepsilon) := \{\boldsymbol{x} \in \Omega \,|\, 0 \le ((h - g - \delta Z) * \varphi_\delta)(\boldsymbol{x}) < \varepsilon \wedge ((g_k - h_k + \delta Z) * \varphi_\delta)(\boldsymbol{x}) \ge 0\}$$

and

$$\bar{W}(\varepsilon) := \{\boldsymbol{x} \in \Omega \,|\, ((h - g - \delta Z) * \varphi_\delta)(\boldsymbol{x}) \ge \varepsilon \wedge ((g_k - h_k + \delta Z) * \varphi_\delta)(\boldsymbol{x}) > 0\},$$

Obviously we have again

(4.31)
$$(\Omega_k \backslash \Omega_\chi) = (\bar{V}(\varepsilon) \cup \bar{W}(\varepsilon)),$$

and

(4.32)
$$\left| \bar{W}(\varepsilon) \right| \le \left| \bar{U}_k(\varepsilon) \right|, \quad \forall \varepsilon > 0.$$

Following the argumentation given above we conclude similarly that $\left| \bar{V}(\varepsilon) \right| \xrightarrow{\varepsilon \to 0} |V(0)| = 0$. If $(\Omega_k \backslash \Omega_\chi) = \emptyset$ then by (4.31), $\bar{V}(\varepsilon) = \emptyset$ for all $\varepsilon \ge 0$. If $(\Omega_k \backslash \Omega_\chi) \ne \emptyset$ then

$\left|\bar{V}_0\right| = \left|\bigcap_{k=1}^{\infty} \bar{V}(\varepsilon^k)\right| = 0$ since $(g - h + \delta Z) * \varphi_\delta \neq 0$ on Ω as shown above. Thus

$$(4.33) \qquad \forall \varepsilon > 0 \, \exists \varepsilon_2 > 0 : \bar{V}(\hat{\varepsilon}) < \frac{\varepsilon}{4}, \quad \forall \hat{\varepsilon} \in (0, \varepsilon_2).$$

Using again Markov's inequality together with (4.26) and (4.27) we analogously obtain

$$(4.34) \qquad \left|\bar{U}_k(\varepsilon)\right| \leq \frac{1}{\varepsilon} \left(\|(g - g_k) * \varphi_\delta\|_{L^1(\Omega)} + \|(h_k - h) * \varphi_\delta\|_{L^1(\Omega)} \right) \overset{k \to \infty}{\longrightarrow} 0, \quad \forall \varepsilon > 0.$$

Hence there exists $K_2 \in \mathbb{N}$ such that

$$(4.35) \qquad \left|\bar{U}_k(\hat{\varepsilon})\right| < \frac{\varepsilon}{4}, \quad \forall k \geq K_2, \forall \hat{\varepsilon} \in (0, \varepsilon_2),$$

with ε_2 as in (4.33). Thus (4.31), (4.32), (4.33) and (4.35) yield

$$(4.36) \qquad |\Omega_k \backslash \Omega_\chi| = \left|\bar{W}(\hat{\varepsilon}) \cup \bar{V}(\hat{\varepsilon})\right| \leq \left|\bar{U}_k(\hat{\varepsilon})\right| + \left|\bar{V}(\hat{\varepsilon})\right| < \frac{\varepsilon}{2}, \quad \forall k \geq K_2, \forall \hat{\varepsilon} \in (0, \varepsilon_2).$$

Hence for $\hat{\varepsilon} \in (0, \min\{\varepsilon_1, \varepsilon_2\})$ and $k \geq \max\{K_1, K_2\}$ (4.30) and (4.36) finally give

$$
\begin{aligned}
|\Omega_\chi \triangle \Omega_k| &= |(\Omega_\chi \backslash \Omega_k) \cup (\Omega_k \backslash \Omega_\chi)| \\
&= |\Omega_\chi \backslash \Omega_k| + |\Omega_k \backslash \Omega_\chi| \\
&\leq |U_k(\hat{\varepsilon})| + |V(\hat{\varepsilon})| + \left|\bar{U}_k(\hat{\varepsilon})\right| + \left|\bar{V}(\hat{\varepsilon})\right| \\
&< \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ was arbitrary this means that $|\Omega_\chi \triangle \Omega_k| \overset{k \to \infty}{\longrightarrow} 0$. $\qquad \square$

Remark 1. *First observe that we proved Lemma 3 for arbitrary $H^1(\Omega)$-sequences $\{E_{b_k}\}_{k \geq 1}$ and $\{E_{f_k}\}_{k \geq 1}$ converging to some limits $E_b^\star$ and $E_f^\star$. Thus the previous result is quite general and does not only hold for back- and foreground images as given by (4.12). Note further that the perturbation $Z \in L^2(\Omega) \backslash H^1(\Omega)$ as well as the analytic mollification appearing in the definitions of Ω_k and Ω_χ from Lemma 3 are purely technical devices. Similarly requiring $\tilde{E}$ to be in $H^1(\Omega)$ instead of $L^2(\Omega)$ (as assumed in Theorem 1) is only necessary for the techniques used in the previous proof. For*

$$\Omega_k = \left\{ \boldsymbol{x} \in \Omega \, \middle| \, \vartheta \left|E_{b_k} - \tilde{E}\right| \leq \left|E_{f_k} - \tilde{E}\right| \right\} \text{ and } \Omega_\chi = \left\{ \boldsymbol{x} \in \Omega \, \middle| \, \vartheta \left|E_b - \tilde{E}\right| \leq \left|E_f - \tilde{E}\right| \right\}$$

it cannot be guaranteed that $\left|E_b - \tilde{E}\right| \neq \left|E_f - \tilde{E}\right|$ on Ω and hence $|V(\varepsilon)| \overset{\varepsilon \to 0}{\nrightarrow} 0$. Thus the definitions of Ω_k and Ω_χ have been modified in order to obtain $|V(\varepsilon)| \overset{\varepsilon \to 0}{\longrightarrow} 0$ and thus the wanted convergence $|\Omega_\chi \triangle \Omega_k| \overset{k \to \infty}{\longrightarrow} 0$.
This may look like a substantial deviation from Algorithm 3.1, however, for any $\alpha \in \mathbb{R}$ we have $\sqrt{|\alpha|^2 + \delta} \overset{\delta \to 0}{\longrightarrow} |\alpha|$, $\delta Z \overset{\delta \to 0}{\longrightarrow} 0$ and for any $g \in L^p(\Omega)$ by Lemma 2 $g * \varphi_\delta \overset{a.e.}{\longrightarrow} g$ as $\delta \to 0$. Hence to satisfy technical necessities we adapt the definition of χ seen in

Algorithm 3.1 in the following way

$$(4.37) \quad \chi(\boldsymbol{x}) := \begin{cases} 1, & \left(\left(\sqrt{\left| E_f - \tilde{E} \right|^2 + \delta} - \vartheta \sqrt{\left| E_b - \tilde{E} \right|^2 + \delta} + \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where we emphasize again that this redefinition is only used for theoretical purposes.

Having analyzed the influence of E_{b} and E_{f} on the edge map χ given by (4.37) we want to prove conversely that there is also a continuous dependence of both back- and foreground images on this edge map. Thus in the following we will use notation that reflects this influence. Let χ be some characteristic function on Ω and

$$(4.38) \qquad\qquad \chi^\delta := \chi * \varphi_\delta,$$

with φ_δ from (4.13). Consider

$$(4.39) \qquad\qquad E[\chi^\delta] := \arg \min_{E \in H^1(\Omega)} J[E, \chi^\delta],$$

and analogously

$$(4.40) \qquad\qquad E[1 - \chi^\delta] := \arg \min_{E \in H^1(\Omega)} J[E, 1 - \chi^\delta].$$

By Lemma 2 the analytic mollification χ^δ satisfies $0 \leq \chi^\delta \leq 1$ since $0 \leq \chi \leq 1$ and thus $\left\| \chi^\delta \right\|_{L^\infty(\Omega)}, \left\| 1 - \chi^\delta \right\|_{L^\infty(\Omega)} \leq 1$. Hence the proof of Theorem 1 shows that the above definitions (4.39) and (4.40) are well posed. In analogy to the discussion in Remark 1 we have to rely on χ^δ instead of χ to be able to prove continuity of back- and foreground images depending on χ. However, this technical adaptation is justifiable as well since Lemma 2 implies $\chi^\delta \xrightarrow{\text{a.e.}} \chi$.

Lemma 4. *Given Assumption 1 and $\tilde{E} \in H^1(\Omega)$ let $\{\chi_k\}_{k \geq 1}$ and χ be characteristic functions satisfying $\chi_k \xrightarrow[L^1(\Omega)]{} \chi$ and denote by χ_k^δ and χ^δ their mollifications in terms of (4.38). Then $E[\chi_k^\delta] \xrightarrow[H^1(\Omega)]{} E[\chi^\delta]$ and $E[1 - \chi_k^\delta] \xrightarrow[H^1(\Omega)]{} E[1 - \chi^\delta]$.*

Proof. We only prove convergence of the sequence $\{E[\chi_k^\delta]\}_{k \geq 1}$. Convergence of $\{E[1 - \chi_k^\delta]\}_{k \geq 1}$ can be shown following the argumentation given below and replacing χ^δ by $1 - \chi^\delta$.

To avoid notational overhead we introduce the shortcuts $E_k := E[\chi_k^\delta]$ and $E_\chi := E[\chi^\delta]$ together with

$$B_k[u, v] := \int_\Omega \beta(\delta + \chi_k^\delta)\nabla u \cdot \nabla v + (\delta + \chi_k^\delta)uv \, d\boldsymbol{x},$$

and

$$f_k[v] := \int_\Omega (\delta + \chi_k^\delta)\tilde{E}v \, d\boldsymbol{x}.$$

Thus E_k is the unique element in $H^1(\Omega)$ satisfying

$$(4.41) \qquad\qquad B_k[E_k, v] = f_k[v], \quad \forall v \in H^1(\Omega).$$

Similarly, E_χ is the unique solution to (4.3) (with χ^δ in place of χ). Combining (4.41) and (4.4) gives

$$(4.42) \qquad B_k[E_\chi, v] = \int_\Omega (\delta + \chi^\delta)\tilde{E}v\,d\boldsymbol{x} + \int_\Omega \beta(\chi_k^\delta - \chi^\delta)\nabla E_\chi \cdot \nabla v + (\chi_k^\delta - \chi^\delta)E_\chi v\,d\boldsymbol{x}.$$

Subtracting (4.42) from (4.41) yields further

$$B_k[E_k - E_\chi, v] = \int_\Omega \tilde{E}v(\chi_k^\delta - \chi^\delta) - \beta(\chi_k^\delta - \chi^\delta)\nabla E_\chi \cdot \nabla v - (\chi_k^\delta - \chi^\delta)E_\chi v\,d\boldsymbol{x},$$

and thus

$$(4.43) \quad |B_k[E_k - E_\chi, v]| \le \left\| \chi_k^\delta - \chi^\delta \right\|_{L^\infty(\Omega)} \|v\|_{H^1(\Omega)} \left(\left\| \tilde{E} \right\|_{H^1(\Omega)} + \|E_\chi\|_{H^1(\Omega)}(\beta + 1) \right).$$

Observe further that due to $\chi_k^\delta \ge 0$, estimate (4.11) shows that all B_k are uniformly coercive

$$(4.44) \qquad \begin{aligned} B_k[v, v] &= \int_\Omega \beta(\delta + \chi_k^\delta)|\nabla v|^2 + (\delta + \chi_k^\delta)v^2 d\boldsymbol{x} \\ &\ge \delta \min\{\beta, 1\} \|v\|_{H^1(\Omega)}^2, \quad \forall v \in H^1(\Omega). \end{aligned}$$

With $p = 1$, $r = \infty$ and $q = \infty$ we conclude by continuity of R_δ from Lemma 2 that $\|\chi_k - \chi\|_{L^1(\Omega)} \overset{k\to\infty}{\longrightarrow} 0$ implies $\left\| \chi_k^\delta - \chi^\delta \right\|_{L^\infty(\Omega)} \overset{k\to\infty}{\longrightarrow} 0$. Hence write $v = E_k - E_\chi$ and combine (4.43) and (4.44) to obtain

$$\|E_k - E_\chi\|_{H^1(\Omega)} \le \frac{1}{\delta \min\{\beta, 1\}} \left\| \chi_k^\delta - \chi^\delta \right\|_{L^\infty(\Omega)} \left(\left\| \tilde{E} \right\|_{H^1(\Omega)} + \|E_\chi\|_{H^1(\Omega)}(\beta + 1) \right) \overset{k\to\infty}{\longrightarrow} 0,$$

thus $\{E_k\}_{k\ge 1}$ converges strongly to E_χ in $H^1(\Omega)$. $\qquad \square$

Combining Lemmas 3 and 4 we can now prove continuity of a map that will be crucial in the following. The mapping T defined below is a functional representation of an adaptation of the steps seen in the Inner Iteration of Algorithm 3.1. Since the set of characteristic functions on Ω do not lend themselves to form a vector space of beneficial mathematical properties we rely on the metric space $\mathcal{M}(\Omega)$ (discussed in Appendix B.2) to define T. Thus we do not study characteristic functions but rather investigate changes of the associated sets. However, Lemma 1 guarantees that "distance" between two sets in the sense of the metric on $\mathcal{M}(\Omega)$ is closely related to the $L^1(\Omega)$-norm difference of the corresponding characteristic functions.

Corollary 2. *Given Assumption 1 and $\tilde{E} \in H^1(\Omega)$ let $\mathcal{M}(\Omega)$ be the metric space defined in Appendix B.2. Further, let $Z \in L^2(\Omega)\backslash H^1(\Omega)$ and φ_δ be given by (4.13). For $\omega \in \mathcal{M}(\Omega)$ let χ_ω denote its characteristic function and χ_ω^δ the mollification in terms of (4.38). Then the mapping $T : \mathcal{M}(\Omega) \to \mathcal{M}(\Omega)$ defined by*

$$T[\omega] := \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\sqrt{\left| E[1 - \chi_\omega^\delta] - \tilde{E} \right|^2 + \delta} - \vartheta\sqrt{\left| E[\chi_\omega^\delta] - \tilde{E} \right|^2 + \delta} + \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) \ge 0 \right. \right\},$$

is continuous.

Proof. Let $\{\omega_k\}_{k \geq 1} \subset \mathcal{M}(\Omega)$ be a convergent sequence with limit ω, i.e., $|\omega \triangle \omega_k| \overset{k \to \infty}{\longrightarrow} 0$, and let $\{\chi_k\}_{k \geq 1}$ and χ respectively be the associated characteristic functions. Then Lemma 1 implies that $\chi_k \underset{L^1(\Omega)}{\longrightarrow} \chi$. Let $E[\chi_k^\delta]$ and $E[1 - \chi_k^\delta]$ denote the images related to χ_k^δ in the sense of (4.39) and (4.40) respectively. Then by Lemma 4 we infer that

$$E[\chi_k^\delta] \underset{H^1(\Omega)}{\longrightarrow} E[\chi^\delta] \text{ and } E[1 - \chi_k^\delta] \underset{H^1(\Omega)}{\longrightarrow} E[1 - \chi^\delta].$$

which according to Lemma 3 yields further $|T[\omega] \triangle T[\omega_k]| \overset{k \to \infty}{\longrightarrow} 0$ and thus $T : \mathcal{M}(\Omega) \to \mathcal{M}(\Omega)$ is continuous. $\qquad \square$

Corollary 2 finally allows us to prove the key result of this section. We want to show that an adapted version of the Inner Iteration of Algorithm 3.1 has a fixed point. We will rely on the classical fixed point theorem by Schauder (see Appendix B.2) for the proof. Thus we have to consider a space that lends itself to some notion of compactness. Hence we extend the mapping T in a way that allows us to consider smoothed edge maps in $L^2(\Omega)$. The map $\mathcal{H}$ below establishes a connection between $L^2(\Omega)$ and $\mathcal{M}(\Omega)$ via thresholding (at $1/2$) and smoothing. For reasons similar to those discussed in Remark 1, a non-smooth function Z is involved as well. The image of some $\xi \in L^2(\Omega)$ under $\mathcal{H}$ is a set and is thus in the domain of T. Hence the introduced functional representation of the Inner Iteration can be applied. Finally a mapping M that assigns a set to its characteristic function together with the operator R_δ that makes for additional smoothing maps back into $L^2(\Omega)$. These modifications allow the use of Schauder's Fixed Point Theorem for the composite map Φ below. Note carefully, however, that these extensions are mere technical devices that are not used in practice. The whole purpose of the function Φ is to show that an adaptation of the Inner Iteration seen in Algorithm 3.1 has a fixed point.

Theorem 2. *Given Assumption 1 and $\tilde{E} \in H^1(\Omega)$ let φ_δ be given by (4.13) and $Z \in L^2(\Omega) \backslash H^1(\Omega)$. For $\xi \in L^2(\Omega)$ define*

$$(4.45) \qquad \qquad \xi^\delta := \xi * \varphi_\delta,$$

and let $\mathcal{H} : L^2(\Omega) \to \mathcal{M}(\Omega)$ be given by

$$(4.46) \qquad \mathcal{H}[\xi] := \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\xi^\delta - \frac{1}{2} + \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) \geq 0 \right. \right\}.$$

Let T be the function defined in Corollary 2 and R_δ the operator from (4.14). Further, let $M : \mathcal{M}(\Omega) \to L^2(\Omega)$ be the mapping that assigns each $\omega \in \mathcal{M}(\Omega)$ to its characteristic function χ_ω, i.e., $M[\omega] = \chi_\omega$.
Then the function

$$(4.47) \qquad \qquad \Phi : \begin{cases} L^2(\Omega) & \to L^2(\Omega), \\ \xi & \mapsto (R_\delta \circ M \circ T \circ \mathcal{H})[\xi], \end{cases}$$

has a fixed point in the set

$$(4.48) \qquad \qquad K := \left\{ \xi \in L^2(\Omega) \left| 0 \leq \xi \leq 1 \text{ a.e. } \Omega \right. \right\}.$$

Proof. We will use Schauder's Fixed Point Theorem (see Appendix B.2) to prove the claim. Thus we show first that the set K is convex in some Banach space X. Choosing $X = L^2(\Omega)$ straightforward calculations show convexity of K: for $\lambda \in (0,1)$ let $\psi := \lambda\xi + (1-\lambda)\bar{\xi}$ be a convex combination of $\xi, \bar{\xi} \in K$. Then obviously $\psi \in L^2(\Omega)$ and since $\lambda \in (0,1)$ implies $1 - \lambda > 0$ and $\xi, \bar{\xi} \in K$ yields $\xi \geq 0$ we see that $\psi \geq 0$ a.e. Ω. Similarly

$$\psi = \lambda\xi + (1-\lambda)\bar{\xi} \leq \lambda + 1 - \lambda = 1, \text{ a.e. } \Omega,$$

since $\xi, \bar{\xi} \in K$. Thus $\psi \in K$ and therefore K is convex.

Next we show continuity of Φ. Since

$$\Phi = R_\delta \circ M \circ T \circ \mathcal{H} : L^2(\Omega) \to \mathcal{M}(\Omega) \to \mathcal{M}(\Omega) \to L^2(\Omega),$$

we start by proving that $\mathcal{H}$ is continuous. Hence let $\{\xi_k\}_{k\geq 1} \subset L^2(\Omega)$ such that $\xi_k \xrightarrow[L^2(\Omega)]{} \xi$ and denote by ξ_k^δ and ξ^δ respectively real analytic mollifications in the sense of (4.45). Further, let $\Omega_k := \mathcal{H}[\xi_k]$ and $\Omega_\chi := \mathcal{H}[\xi]$ and let $\{\chi_k\}_{k\geq 1}$ and χ respectively denote the corresponding characteristic functions. We want to show that $|\Omega_\chi \triangle \Omega_k| \xrightarrow{k\to\infty} 0$. We compute

(4.49)
$$\Omega_\chi \backslash \Omega_k = \{ \boldsymbol{x} \in \Omega \,| \chi(\boldsymbol{x}) = 1 \wedge \chi_k(\boldsymbol{x}) = 0 \}$$
$$= \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\xi^\delta - \frac{1}{2} + \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) \geq 0 \wedge \right.\right.$$
$$\left.\left(\left(\frac{1}{2} - \xi_k^\delta - \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) > 0 \right\}$$

For $\varepsilon > 0$ we introduce the set

$$U_k(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega | \left((\xi^\delta - \xi_k^\delta) * \varphi_\delta \right)(\boldsymbol{x}) \geq \varepsilon \right\}.$$

By distributivity of the convolution it follows immediately that $(\Omega_\chi \backslash \Omega_k) \subset U_k(0)$. Using a similar argumentation as in the proof of Lemma 3 we start by defining

$$V(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega \left| 0 \leq \left(\left(\xi^\delta - \frac{1}{2} + \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) < \varepsilon \wedge \left(\left(\frac{1}{2} - \xi_k^\delta - \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) > 0 \right\},\right.$$

and

$$W(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\xi^\delta - \frac{1}{2} + \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) \geq \varepsilon \wedge \left(\left(\frac{1}{2} - \xi_k^\delta - \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) > 0 \right\},\right.$$

so that we may rewrite (4.49) in the following manner

(4.50) $$(\Omega_\chi \backslash \Omega_k) = (V(\varepsilon) \cup W(\varepsilon)).$$

Obviously

(4.51) $$|W(\varepsilon)| \leq |U_k(\varepsilon)|, \quad \forall \varepsilon > 0,$$

and we will show that $|V(\varepsilon)| \to 0$ as $\varepsilon \to 0$. If $(\Omega_\chi \backslash \Omega_k) = \emptyset$ then by (4.50), $V(\varepsilon) = \emptyset$ for

all $\varepsilon \geq 0$ which makes the claim trivial. Thus suppose $(\Omega_\chi \backslash \Omega_k) \neq \emptyset$ and let $\{\varepsilon^k\}_{k \geq 1}$ denote a monotonically decreasing null sequence. Then again the definition of $V(\varepsilon)$ implies that $\{V(\varepsilon^k)\}_{k \geq 1}$ is a decreasing nested sequence of sets such that $V_0 := \cap_{k=1}^\infty V(\varepsilon^k)$. Assume $|V_0| > 0$ then

$$(4.52) \qquad (\xi^\delta - \frac{1}{2} + \delta Z) * \varphi_\delta = 0 \text{ on } V_0.$$

Mimicking the argumentation given in the proof of Lemma 3 we conclude by Lemma 2 that $(\xi^\delta - \frac{1}{2} + \delta Z) * \varphi_\delta$ is a real analytic expression on Ω thus (4.52) has to hold on the whole image domain Ω. Stating this in terms of the operator R_δ from (4.14) and exploiting the fact that $R_\delta[0] = 0$ we obtain analogously

$$0 = R_\delta[\xi^\delta - \frac{1}{2} + \delta Z] = R_\delta[0],$$

and thus by injectivity of R_δ (compare Lemma 2)

$$\xi^\delta - \frac{1}{2} + \delta Z = 0 \text{ on } \Omega,$$

or equivalently

$$\delta Z = \frac{1}{2} - \xi^\delta \text{ on } \Omega.$$

This is a contradiction since ξ^δ is real analytic (by Lemma 2) but Z is by assumption in $L^2(\Omega) \backslash H^1(\Omega)$. Thus V_0 cannot have positive measure and $|V(\varepsilon)| \xrightarrow{\varepsilon \to 0} 0$. Hence

$$(4.53) \qquad \forall \varepsilon > 0 \, \exists \varepsilon_1 > 0 : V(\hat{\varepsilon}) < \frac{\varepsilon}{4}, \quad \forall \hat{\varepsilon} \in (0, \varepsilon_1).$$

Using (4.45) together with Chebyshev's inequality and Young's inequality for convolutions (both in Appendix B.1) we obtain

$$
\begin{aligned}
|U_k(\varepsilon)| \leq & \frac{1}{\varepsilon^2} \int_\Omega \left| (\xi^\delta - \xi_k^\delta) * \varphi_\delta \right|^2 d\boldsymbol{x} \\
= & \frac{1}{\varepsilon^2} \left\| (\xi^\delta - \xi_k^\delta) * \varphi_\delta \right\|_{L^2(\Omega)}^2 \\
\leq & \frac{1}{\varepsilon^2} \left\| \xi^\delta - \xi_k^\delta \right\|_{L^2(\Omega)}^2 \|\varphi_\delta\|_{L^1(\Omega)}^2 \\
= & \frac{1}{\varepsilon^2} \|(\xi - \xi_k) * \varphi_\delta\|_{L^2(\Omega)}^2 \|\varphi_\delta\|_{L^1(\Omega)}^2 \\
\leq & \frac{1}{\varepsilon^2} \|\xi - \xi_k\|_{L^2(\Omega)}^2 \|\varphi_\delta\|_{L^1(\Omega)}^4 \xrightarrow{k \to \infty} 0, \quad \forall \varepsilon > 0.
\end{aligned}
$$

Thus we infer that for $\varepsilon_1 > 0$ appearing in (4.53) there exists an index $K_1 \in \mathbb{N}$ such that

$$(4.54) \qquad |U_k(\hat{\varepsilon})| < \frac{\varepsilon}{4}, \quad \forall k \geq K_1, \forall \hat{\varepsilon} \in (0, \varepsilon_1).$$

Conversely for

$$\Omega_k \backslash \Omega_\chi = \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\frac{1}{2} - \xi^\delta - \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) > 0 \wedge \left(\left(\xi_k^\delta - \frac{1}{2} + \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) \geq 0 \right\} \right.,$$

we define

$$\bar{U}_k(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega \left| \left((\xi_k^\delta - \xi^\delta) * \varphi_\delta \right) (\boldsymbol{x}) \geq \varepsilon \right\} \right.,$$

and we get similarly $(\Omega_k \backslash \Omega_\chi) \subset \bar{U}_k(0)$. We have again

(4.55) $$(\Omega_k \backslash \Omega_\chi) = (\bar{V}(\varepsilon) \cup \bar{W}(\varepsilon)),$$

for

$$\bar{V}(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega \left| 0 \leq \left(\left(\frac{1}{2} - \xi^\delta - \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) < \varepsilon \wedge \left(\left(\xi_k^\delta - \frac{1}{2} + \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) \geq 0 \right\} \right.,$$

and

$$\bar{W}(\varepsilon) := \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\frac{1}{2} - \xi^\delta - \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) \geq \varepsilon \wedge \left(\left(\xi_k^\delta - \frac{1}{2} + \delta Z \right) * \varphi_\delta \right) (\boldsymbol{x}) \geq 0 \right\} \right.,$$

such that

(4.56) $$\left| \bar{W}(\varepsilon) \right| \leq \left| \bar{U}_k(\varepsilon) \right|, \quad \forall \varepsilon > 0.$$

Following the argumentation given above we conclude similarly that $\left| \bar{V}(\varepsilon) \right| \xrightarrow{\varepsilon \to 0} 0$: if $(\Omega_k \backslash \Omega_\chi) = \emptyset$ then $\bar{V}(\varepsilon) = \emptyset$ for any $\varepsilon \geq 0$. If $(\Omega_k \backslash \Omega_\chi) \neq \emptyset$ then $\left| \bar{V}_0 \right| = \left| \bigcap_{k=1}^\infty \bar{V}(\varepsilon^k) \right| = 0$ since $(\xi^\delta - \frac{1}{2} + \delta Z) * \varphi_\delta \neq 0$ on Ω as shown above. Thus

(4.57) $$\forall \varepsilon > 0 \, \exists \varepsilon_2 > 0 : \bar{V}(\hat{\varepsilon}) < \frac{\varepsilon}{4}, \quad \forall \hat{\varepsilon} \in (0, \varepsilon_2).$$

Using again (4.45) together with Chebyshev's inequality and Young's inequality for convolutions we similarly obtain

$$\left| \bar{U}_k(\varepsilon) \right| \leq \frac{1}{\varepsilon^2} \| \xi - \xi_k \|_{L^2(\Omega)}^2 \| \varphi_\delta \|_{L^1(\Omega)}^4 \xrightarrow{k \to \infty} 0, \quad \forall \varepsilon > 0,$$

and hence deduce the existence of $K_2 \in \mathbb{N}$ such that

(4.58) $$\left| \bar{U}_k(\hat{\varepsilon}) \right| < \frac{\varepsilon}{4}, \quad \forall k \geq K_2, \forall \hat{\varepsilon} \in (0, \varepsilon_2),$$

with $\varepsilon_2 > 0$ as in (4.57). Thus for $\hat{\varepsilon} \in (0, \min\{\varepsilon_1, \varepsilon_2\})$ and $k \geq \max\{K_1, K_2\}$ we combine (4.50), (4.51) and (4.53)–(4.58) to obtain

(4.59)
$$\begin{aligned} |\mathcal{H}[\xi] \, \triangle \, \mathcal{H}[\xi_k]| &= |\Omega_\chi \, \triangle \, \Omega_k| \\ &= |(\Omega_\chi \backslash \Omega_k) \cup (\Omega_k \backslash \Omega_\chi)| \\ &= |\Omega_\chi \backslash \Omega_k| + |\Omega_k \backslash \Omega_\chi| \\ &\leq |U_k(\hat{\varepsilon})| + |V(\hat{\varepsilon})| + \left| \bar{U}_k(\hat{\varepsilon}) \right| + \left| \bar{V}(\hat{\varepsilon}) \right| \\ &< \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary this means that $|\mathcal{H}[\xi] \, \triangle \, \mathcal{H}[\xi_k]| \xrightarrow{k \to \infty} 0$ and hence $\mathcal{H}$ is continuous. According to Corollary 2, $T : \mathcal{M}(\Omega) \to \mathcal{M}(\Omega)$ is continuous as well. With $p = 2$, $r = 2$ and $q = 1$ Lemma 2 implies that R_δ is also continuous. Continuity of the map $M : \mathcal{M}(\Omega) \to L^2(\Omega)$ is shown as follows. Hence let $\{\omega_k\}_{k \geq 1} \subset \mathcal{M}(\Omega)$ and $\omega \in \mathcal{M}(\Omega)$ such that $|\omega_k \, \triangle \, \omega| \xrightarrow{k \to \infty} 0$. Then by Lemma 1 the associated characteristic functions $\{\chi_k\}_{k \geq 1}$ and χ_ω respectively satisfy $\chi_k \xrightarrow[L^1(\Omega)]{} \chi_\omega$ and hence

$$\|\chi_k - \chi_\omega\|^2_{L^2(\Omega)} = \int_\Omega |\chi_k - \chi_\omega|^2 \, d\boldsymbol{x} = \int_\Omega |\chi_k - \chi_\omega| \, d\boldsymbol{x} = \|\chi_k - \chi_\omega\|_{L^1(\Omega)} \xrightarrow{k \to \infty} 0,$$

thus M is continuous. Hence $\Phi = R_\delta \circ M \circ T \circ \mathcal{H} : L^2(\Omega) \to L^2(\Omega)$ is continuous. Left to show is that $\Phi(K) \subseteq K$ and $\Phi(K)$ is compact in K. Therefore, note that Lemma 2 implies $0 \leq R_\delta[\chi_\omega] \leq 1$ for any characteristic function χ_ω and further $R_\delta[\chi_\omega] = \chi_\omega * \varphi_\delta \in H^1(\Omega) \subset L^2(\Omega)$. Thus $R_\delta[\chi_\omega] \in K$ and indeed $\Phi(K) \subseteq K$. Now choose $\bar{\xi} \in K$ arbitrary and let $\mathcal{H}[\bar{\xi}] = \bar{\omega}$. Let further $T[\bar{\omega}] = \omega$, denote its characteristic function by χ_ω, i.e., $M[\omega] = \chi_\omega$, and set $\chi_\omega^\delta = R_\delta[\chi_\omega] = \varphi_\delta * \chi_\omega$. Hence $\chi_\omega^\delta = \Phi[\bar{\xi}]$. Note that (compare [Eva08, App. C.4, Th. 6])

$$\left| \nabla \chi_\omega^\delta(\boldsymbol{x}) \right| = \left| \int_\Omega \nabla \varphi_\delta(\boldsymbol{x} - \boldsymbol{y}) \chi_\omega(\boldsymbol{y}) \, d\boldsymbol{y} \right|$$
$$\leq \|\chi_\omega\|_{L^\infty(\Omega)} \int_\Omega |\nabla \varphi_\delta(\boldsymbol{x} - \boldsymbol{y})| \, d\boldsymbol{y}$$
$$\leq \|1\|_{L^2(\Omega)} \|\nabla \varphi_\delta\|_{L^2(\Omega)}$$
$$= \sqrt{|\Omega|} \, \|\nabla \varphi_\delta\|_{L^2(\Omega)},$$

and thus

$$\left\| \nabla \chi_\omega^\delta \right\|^2_{L^2(\Omega)} = \int_\Omega \left| \int_\Omega \nabla \varphi_\delta(\boldsymbol{x} - \boldsymbol{y}) \chi_\omega(\boldsymbol{y}) \, d\boldsymbol{y} \right|^2 d\boldsymbol{x}$$
$$\leq \int_\Omega |\Omega| \, \|\nabla \varphi_\delta\|^2_{L^2(\Omega)} \, d\boldsymbol{x}$$
$$= |\Omega|^2 \, \|\nabla \varphi_\delta\|^2_{L^2(\Omega)}.$$

Since $\chi_\omega^\delta \in H^1(\Omega)$ we compute further (using again Young's inequality for convolutions)

$$(4.60) \qquad \begin{aligned} \left\| \Phi[\bar{\xi}] \right\|^2_{H^1(\Omega)} &= \left\| \chi_\omega^\delta \right\|^2_{H^1(\Omega)} = \left\| \chi_\omega^\delta \right\|^2_{L^2(\Omega)} + \left\| \nabla \chi_\omega^\delta \right\|^2_{L^2(\Omega)} \\ &\leq \|\chi_\omega * \varphi_\delta\|^2_{L^2(\Omega)} + |\Omega|^2 \, \|\nabla \varphi_\delta\|^2_{L^2(\Omega)} \\ &\leq \|\chi_\omega\|^2_{L^1(\Omega)} \|\varphi_\delta\|^2_{L^2(\Omega)} + |\Omega|^2 \, \|\nabla \varphi_\delta\|^2_{L^2(\Omega)} \\ &\leq \|1\|^2_{L^1(\Omega)} \|\varphi_\delta\|^2_{L^2(\Omega)} + |\Omega|^2 \, \|\nabla \varphi_\delta\|^2_{L^2(\Omega)} \\ &= |\Omega|^2 \left(\|\varphi_\delta\|^2_{L^2(\Omega)} + \|\nabla \varphi_\delta\|^2_{L^2(\Omega)} \right) =: c(\delta). \end{aligned}$$

Thus any sequence $\{\Phi[\xi_k]\}_{k \geq 1} \subset K \subset H^1(\Omega)$ is bounded in the $H^1(\Omega)$-norm in terms of fixed $c(\delta)$. Since Ω is bounded $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$ [Ada75, Chap. 6, Th. 6.53] so $\{\Phi[\xi_k]\}_{k \geq 1}$ has a convergent subsequence. Thus $\Phi(K)$ is compact in

K and we finally conclude by Schauder's Fixed Point Theorem that Φ has a fixed point in K. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

In the next section we show a result similar to Theorem 1 addressing existence and uniqueness of the smooth image I_{s} as seen in the Outer Iteration of Algorithm 3.1. Furthermore, in analogy to the map T from Corollary 2 we introduce a continuous representation of the Outer Iteration.

4.2 The Outer Iteration

Given Assumption 1 let χ be a characteristic function and $I \in H^2(\Omega)$. Recall the functional F given by (3.1). We want to deduce the necessary optimality conditions for the minimization of F with respect to I for a fixed χ. To simplify notation we define

$$(4.61) \qquad\qquad F_{\mathrm{s}}[I] := F[I, \chi],$$

and consider the minimization problem

$$(4.62) \qquad\qquad \inf_{I \in H^2(\Omega)} F_{\mathrm{s}}[I].$$

We compute the Gâteaux derivative (see Appendix B.2) of F_{s} in an arbitrary direction $v \in C^\infty(\bar{\Omega})$

$$
\begin{aligned}
\frac{\delta F_{\mathrm{s}}}{\delta I}[I; v] &= \frac{d}{dt}\left(\frac{1}{2}\int_\Omega \left|I + tv - \tilde{I}\right|^2 (\delta + \chi) + \beta(\delta + \chi)\left|\nabla^2(I + tv)\right|^2 d\boldsymbol{x}\right)\bigg|_{t=0} \\
&= \int_\Omega (I + tv - \tilde{I})(\delta + \chi)v + \beta(\delta + \chi)\nabla^2(I + tv) : \nabla^2 v\, d\boldsymbol{x}\bigg|_{t=0} \\
&= \int_\Omega (I - \tilde{I})(\delta + \chi)v + \beta(\delta + \chi)\nabla^2 I : \nabla^2 v\, d\boldsymbol{x},
\end{aligned}
$$

where ":" denotes a component-wise matrix scalar product (sometimes called Frobenius inner product, see Appendix A) such that $\left|\nabla^2 v\right|^2 = \nabla^2 v : \nabla^2 v$. Thus the weak formulation of the necessary optimality conditions for (4.62) is given by

$$(4.63) \qquad\qquad A_{\mathrm{s}}[u, v] = d_{\mathrm{s}}[v], \quad \forall v \in H^2(\Omega),$$

where $A_{\mathrm{s}} : H^2(\Omega) \times H^2(\Omega) \to \mathbb{R}$ is a bilinear form defined by

$$(4.64) \qquad\qquad A_{\mathrm{s}}[u, v] := \int_\Omega \beta(\delta + \chi)\nabla^2 u : \nabla^2 v + (\delta + \chi)uv\, d\boldsymbol{x},$$

and $d_{\mathrm{s}} : H^2(\Omega) \to \mathbb{R}$ is a linear functional

$$(4.65) \qquad\qquad d_{\mathrm{s}}[v] := \int_\Omega \tilde{I}(\delta + \chi)v\, d\boldsymbol{x}.$$

We start by showing that the minimization problem (4.62) has a unique solution.

Theorem 3. *Given Assumption 1 let χ be a characteristic function on Ω. Then the minimization problem (4.62) has a unique solution.*

Proof. We want to apply the Lax–Milgram Lemma (Appendix B.2). Thus by employing $\|\chi\|_{L^\infty(\Omega)} \leq 1$ we see that d_{s} is bounded

$$|d_{\mathrm{s}}[v]| \leq \|\delta + \chi\|_{L^\infty(\Omega)} \int_\Omega \left|\tilde{I}v\right| d\boldsymbol{x} \leq (\delta + 1) \left\|\tilde{I}\right\|_{L^2(\Omega)} \|v\|_{H^2(\Omega)}.$$

Next we show that the bilinear form A_{s} is bounded as well

$$|A_{\mathrm{s}}[u,v]| \leq \|\beta(\delta + \chi)\|_{L^\infty(\Omega)} \int_\Omega \left|\nabla^2 u\right| \left|\nabla^2 v\right| d\boldsymbol{x} + \|\delta + \chi\|_{L^\infty(\Omega)} \int_\Omega |u||v| d\boldsymbol{x}$$
$$\leq \beta(\delta + \|\chi\|_{L^\infty(\Omega)}) \left\|\nabla^2 u\right\|_{L^2(\Omega)} \left\|\nabla^2 u\right\|_{L^2(\Omega)} + (\delta + \|\chi\|_{L^\infty(\Omega)}) \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)}$$
$$\leq (\delta + 1)(\beta + 1) \|u\|_{H^2(\Omega)} \|v\|_{H^2(\Omega)},$$

where we used the Cauchy–Schwarz inequality (Appendix B.1). Left to show is coercivity of A_{s}. We compute

(4.66)
$$A_{\mathrm{s}}[u,u] = \int_\Omega \beta \left(\delta + \chi\right) \nabla^2 u : \nabla^2 u + (\delta + \chi) u^2 d\boldsymbol{x}$$
$$\geq \delta \int_\Omega \beta \left|\nabla^2 u\right|^2 + u^2 d\boldsymbol{x}$$
$$\geq \delta \min\{\beta, 1\} \left(\left\|\nabla^2 u\right\|_{L^2(\Omega)}^2 + \|u\|_{L^2(\Omega)}^2\right).$$

By Corollary 4.16 in [Ada75] the right hand side of (4.66) is equivalent to $\|\cdot\|_{H^2(\Omega)}$ so coercivity follows. Thus the Lax–Milgram Lemma implies well posedness of (4.63). This together with the fact that the cost functional F_{s} is strictly convex in I ensures that the minimization problem (4.62) has a unique solution [ET99, Chap. 2, Proposition 1.2]. $\quad\square$

Thus Theorem 3 guarantees well posedness of the following definition

(4.67)
$$I_{\mathrm{s}} := \arg \min_{I \in H^2(\Omega)} F_{\mathrm{s}}[I],$$

with F_{s} as given by (4.61). Similar to the analysis done for the Inner Iteration in the previous section we will now investigate the influence of the underlying edge maps on the quantities computed in the Outer Iteration. Thus let χ be some characteristic function on Ω and χ^δ its analytic mollification in the sense of (4.38). In analogy to (4.39) we introduce the notation

(4.68)
$$I[\chi^\delta] := \arg \min_{I \in H^2(\Omega)} F[I, \chi^\delta].$$

Since by Lemma 2, $0 \leq \chi^\delta \leq 1$ holds, the proof of Theorem 3 shows that (4.68) is well defined. As was the case for Lemma 4 we have to rely here again on χ^δ instead of χ to be able to show continuity of the smooth image with respect to χ. But again Lemma 2 yields $\chi^\delta \xrightarrow{\mathrm{a.e.}} \chi$ which implies that the deviation of (4.68) to the smooth image used in practice is negligible for δ vanishingly small.

Lemma 5. *Given Assumption 1 let $\{\chi_k\}_{k\geq 1}$ and χ be characteristic functions satisfying $\chi_k \underset{L^1(\Omega)}{\longrightarrow} \chi$. Then $I[\chi_k^\delta] \underset{H^2(\Omega)}{\longrightarrow} I[\chi^\delta]$ with χ_k^δ and χ^δ denoting the analytic mollifications of χ_k and χ respectively in the sense of (4.38).*

Proof. For the sake of brevity let $I_k := I[\chi_k^\delta]$ and $I_\chi := I[\chi^\delta]$. Then by Theorem 3, I_k uniquely satisfies

$$A_k[I_k, v] = d_k[v], \quad \forall v \in H^2(\Omega),$$

with

$$(4.69) \qquad A_k[u, v] := \int_\Omega \beta(\delta + \chi_k^\delta)\nabla^2 u : \nabla^2 v + (\delta + \chi_k^\delta)uv \, d\boldsymbol{x},$$

and

$$d_k[v] := \int_\Omega (\delta + \chi_k^\delta)\tilde{I}v \, d\boldsymbol{x}.$$

Similarly Theorem 3 implies that I_χ is the unique solution to (4.63) with χ^δ in place of χ. Combining (4.63) and (4.69) yields

$$(4.70) \qquad A_k[I_\chi, v] = \int_\Omega (\delta + \chi^\delta)\tilde{I}v \, d\boldsymbol{x} + \int_\Omega \beta(\chi_k^\delta - \chi^\delta)\nabla^2 I_\chi : \nabla^2 v + (\chi_k^\delta - \chi^\delta)I_\chi v \, d\boldsymbol{x}.$$

Subtracting (4.70) from (4.69) gives

$$(4.71) \qquad \begin{aligned} A_k[I_k - I_\chi, v] = &\int_\Omega \tilde{I}v(\chi_k^\delta - \chi^\delta) - \beta(\chi_k^\delta - \chi^\delta)\nabla^2 I_\chi : \nabla^2 v \, d\boldsymbol{x} \\ &- \int_\Omega (\chi_k^\delta - \chi^\delta)I_\chi v \, d\boldsymbol{x}, \end{aligned}$$

and thus

$$(4.72) \qquad |A_k[I_k - I_\chi, v]| \leq \left\|\chi_k^\delta - \chi^\delta\right\|_{L^\infty(\Omega)} \|v\|_{H^2(\Omega)} \left(\left\|\tilde{I}\right\|_{L^2(\Omega)} + \|I_\chi\|_{H^2(\Omega)}(\beta + 1)\right).$$

According to the proof of Theorem 3, A_k is coercive and thus (see Appendix B.2 for details)

$$(4.73) \qquad c_k := \inf_{v \in H^2(\Omega)} \frac{|A_k[v, v]|}{\|v\|_{H^2(\Omega)}^2} > 0, \quad \forall k \in \mathbb{N}.$$

By coercivity of A_{s} we obtain similarly that its coercivity constant c_{s} satisfies $c_{\mathrm{s}} > 0$. With $p = 1$, $r = \infty$ and $q = \infty$ we conclude by continuity of R_δ from Lemma 2 that $\|\chi_k - \chi\|_{L^1(\Omega)} \overset{k\to\infty}{\longrightarrow} 0$ implies

$$(4.74) \qquad \left\|\chi_k^\delta - \chi^\delta\right\|_{L^\infty(\Omega)} \overset{k\to\infty}{\longrightarrow} 0.$$

Hence we have for all $v \in H^2(\Omega)$

$$|A_k[v,v] - A[v,v]| = \left| \int_\Omega \beta(\chi_k^\delta - \chi^\delta) \left|\nabla^2 v\right|^2 + (\chi_k^\delta - \chi^\delta)v^2 d\boldsymbol{x} \right|$$

$$\leq \left\|\chi_k^\delta - \chi^\delta\right\|_{L^\infty(\Omega)} \|v\|_{H^2(\Omega)}^2 (\beta + 1) \overset{k\to\infty}{\longrightarrow} 0,$$

and thus by adding and subtracting $A[v,v]$ in (4.73) we conclude that $c_k \overset{k\to\infty}{\longrightarrow} c_\mathrm{s}$. Since $c_\mathrm{s} > 0$ there exists $\eta > 0$ such that $c_k > \eta > 0$ for all k large enough. Hence for these k we get

(4.75)
$$|A_k[v,v]| \geq c_k \|v\|_{H^2(\Omega)}^2 > \eta \|v\|_{H^2(\Omega)}^2, \quad \forall v \in H^2(\Omega).$$

Setting $v = I_k - I_\chi$ we combine (4.72), (4.74) and (4.75) to get

$$\|I_k - I_\chi\|_{H^2(\Omega)} \leq \frac{1}{\eta} \left\|\chi_k^\delta - \chi^\delta\right\|_{L^\infty(\Omega)} \left(\left\|\tilde{I}\right\|_{L^2(\Omega)} + (\beta + 1) \|I_\chi\|_{H^2(\Omega)}\right) \overset{k\to\infty}{\longrightarrow} 0,$$

thus $I_k \underset{H^2(\Omega)}{\longrightarrow} I_\chi$. $\qquad\qquad\square$

We now want to show that there is also a continuous dependence of the fuzzy edge map on the smooth image $I[\chi^\delta]$ as given by (4.68). However, we make an important observation first. In Algorithm 3.1 the fuzzy edge map $\tilde{E}$ links the Outer to the Inner Iteration: the smooth image I_s determines $\tilde{E}$ which is then segmented in the Inner Iteration. Observe further that we have made no specific assumptions on $\tilde{E}$ in the previous section except it being in $H^1(\Omega)$. (Theorem 1 was proven for $\tilde{E} \in L^2(\Omega)$ to clarify that well posedness of E_b and E_f only requires standard regularity of the right hand side.) Thus in order to exploit the results obtained for the Inner Iteration we have to ensure that $\tilde{E} \in H^1(\Omega)$. However, Note that $\tilde{E}$ as used in Algorithm 3.1 only has a weak derivative where $|\nabla I_\mathrm{s}| > 0$, which cannot be guaranteed on the whole image domain Ω. Moreover, we want to utilize continuity of $I[\chi^\delta]$ with respect to χ. Hence we introduce the following modification of $\tilde{E}$

(4.76)
$$\tilde{E} := \left|\nabla I_\mathrm{s}^\delta + \delta\right|,$$

where $I_s^\delta := I[\chi^\delta]$. Since $I_\mathrm{s}^\delta \in H^2(\Omega)$ we get

$$\int_\Omega \tilde{E}^2 d\boldsymbol{x} = \int_\Omega \left|\nabla I_\mathrm{s}^\delta + \delta\right|^2 d\boldsymbol{x} < \infty,$$

and

$$\nabla \tilde{E} = \frac{\nabla^2 I_\mathrm{s}^\delta \nabla I_\mathrm{s}^\delta}{|\nabla I_\mathrm{s}^\delta + \delta|} \in L^2(\Omega),$$

thus indeed $\tilde{E} \in H^1(\Omega)$. By Assumption 1, δ is small; thus, the deviation of (4.76) to the fuzzy edge map used in practice is negligible. Moreover, note that $\tilde{E} \in H^1(\Omega)$ was only necessary to rule out the possibility that $|V_0| = 0$ in the proof of Lemma 3 (compare Remark 1). Thus (4.76) just like the redefinition (4.37) of χ is only a technical device to satisfy theoretical requirements. Now we prove the following general result.

Lemma 6. *Given Assumption 1 the mapping*

$$(4.77) \qquad \mathcal{I} : \begin{cases} H^2(\Omega) & \to H^1(\Omega), \\ I & \mapsto |\nabla I + \delta|, \end{cases}$$

is continuous.

Proof. Observe first that $\mathcal{I}$ is well defined since for $I \in H^2(\Omega)$ the image $\mathcal{I}[I]$ is indeed in $H^1(\Omega)$ as detailed above. Let $\{I_k\}_{k \geq 1}$ be a sequence in $H^2(\Omega)$ strongly converging to $I \in H^2(\Omega)$ and define $\tilde{E}_k := |\nabla I_k + \delta|$ and similarly $\tilde{E} := |\nabla I + \delta|$. Then by the reverse triangle inequality we obtain

$$\left\| \tilde{E}_k - \tilde{E} \right\|_{H^1(\Omega)} = \left\| |\nabla I_k + \delta| - |\nabla I + \delta| \right\|_{H^1(\Omega)}$$
$$\leq \|\nabla I_k - \nabla I\|_{H^1(\Omega)}$$
$$\leq \|I_k - I\|_{H^2(\Omega)} \overset{k \to \infty}{\longrightarrow} 0,$$

thus $\mathcal{I}$ is continuous. $\qquad\qquad\square$

Observe carefully that, similar to the situation in Section 4.1, we are now considering an adapted version of the Outer Iteration from Algorithm 3.1. As explained above the changes introduced in the computations of $\tilde{E}$ and $I[\chi^\delta]$ are negligible for vanishingly small δ. However, these slight modifications allow us to prove crucial results in the following. Note further that we can write this modified Outer Iteration as a composite function evaluation in the following form. For a given characteristic function χ let

$$(4.78) \qquad \tilde{E} = \mathcal{I}\left[I[R_\delta[\chi]] \right],$$

where R_δ given by (4.14), $I[\cdot]$ is defined in (4.68) and $\mathcal{I}$ is given in (4.77). In the following section we establish the connection of (4.78) to the adapted Inner Iteration discussed in Section 4.1.

4.3 Rewriting the Algorithm

We will now investigate the dependence of the adapted Inner Iteration from Section 4.1 on the fuzzy edge map $\tilde{E}$ given by the modified Outer Iteration (4.78). Therefore it is necessary to clarify the several dependencies on $\tilde{E}$ that appear in the Inner Iteration. Observe first that $\tilde{E}$ appears in the linear functional that determines the right hand sides f_b and f_f of (4.3) and (4.8) respectively. Thus E_b and E_f do not only depend on χ but also on $\tilde{E}$. Note further that in the introduced functional representation of the Inner Iteration given by the mapping T from Corollary 2 we did not consider any dependence on $\tilde{E}$. We only proved continuity of T for a fixed $\tilde{E}$. However, in general view of Algorithm 3.1, sets given by T depend implicitly on the fuzzy edge map $\tilde{E}$. To reflect this double dependence we introduce notation similar to (4.39). Let $E[\chi^\delta, \tilde{E}]$ denote the background image for a given mollified edge map χ^δ as defined in (4.38) and a certain fuzzy edge map $\tilde{E}$ as given by (4.76). Similarly we denote a foreground image for given χ^δ and $\tilde{E}$ by $E[1 - \chi^\delta, \tilde{E}]$. Using this notation we can prove the following extension of Corollary 2.

Corollary 3. *Given Assumption 1 let φ_δ be defined by (4.13), $Z \in L^2(\Omega) \backslash H^1(\Omega)$ and $\tilde{E} \in H^1(\Omega)$. For $\omega \in \mathcal{M}(\Omega)$ let χ_ω denote its characteristic function and $\chi_\omega^\delta = R_\delta[\chi_\omega]$. Then the mapping $\mathcal{T} : \mathcal{M}(\Omega) \times H^1(\Omega) \to \mathcal{M}(\Omega)$ defined by*

$$\mathcal{T}[\omega, \tilde{E}] := \left\{ \boldsymbol{x} \in \Omega \left| \left(\left(\sqrt{\left| E[1 - \chi_\omega^\delta, \tilde{E}] - \tilde{E} \right|^2 + \delta} \right. \right. \right. \right.$$
$$\left. \left. \left. \left. - \vartheta \sqrt{\left| E[\chi_\omega^\delta, \tilde{E}] - \tilde{E} \right|^2 + \delta} + \delta Z \right) * \varphi_\delta \right)(\boldsymbol{x}) \geq 0 \right\},$$

is continuous.

Proof. Let $\{\tilde{E}_k\}_{k \geq 1} \subset H^1(\Omega)$ be a sequence strongly converging to $\tilde{E} \in H^1(\Omega)$ and let $\{\omega_k\}_{k \geq 1}$ be a sequence in $\mathcal{M}(\Omega)$ converging to ω. Let further χ_k and χ_ω denote the characteristic functions associated to ω_k and ω respectively. Thus by Lemma 1 we conclude $\chi_k \underset{L^1(\Omega)}{\longrightarrow} \chi_\omega$. For the sake of brevity let $E_{b_k} := E[\chi_k^\delta, \tilde{E}_k]$, $E_b := E[\chi_\omega^\delta, \tilde{E}]$, $E_{f_k} := E[1 - \chi_k^\delta, \tilde{E}_k]$ and $E_f := E[1 - \chi_\omega^\delta, \tilde{E}]$. We prove that $E_{b_k} \underset{H^1(\Omega)}{\longrightarrow} E_b$. Analogous to the proof of Lemma 4, convergence of $\{E_{f_k}\}_{k \geq 1}$ follows similarly by replacing χ_ω^δ with $1 - \chi_\omega^\delta$.

Theorem 1 implies that E_{b_k} is the unique solution to the weak problem

$$(4.79) \qquad\qquad B_k[u, v] = f_k[v], \quad \forall v \in H^1(\Omega),$$

with

$$B_k[u, v] := \int_\Omega (\delta + \chi_k^\delta) \nabla u \cdot \nabla v + (\delta + \chi_k^\delta) uv \, d\boldsymbol{x},$$

and

$$f_k[v] := \int_\Omega (\delta + \chi_k^\delta) \tilde{E}_k v \, d\boldsymbol{x}.$$

Similarly E_b uniquely satisfies (4.3) (with χ_ω^δ in place of χ). Combining (4.79) and (4.3) hence gives

$$B_k[E_{b_k} - E_b, v] = \int_\Omega \delta v(\tilde{E}_k - \tilde{E}) + v(\tilde{E}_k \chi_k^\delta - \tilde{E} \chi_\omega^\delta) \, d\boldsymbol{x}$$
$$- \int_\Omega \beta(\chi_k^\delta - \chi_\omega^\delta) \nabla E_b \cdot \nabla v + (\chi_k^\delta - \chi_\omega^\delta) E_b v \, d\boldsymbol{x}$$
$$= \int_\Omega \delta v(\tilde{E}_k - \tilde{E}) + v\tilde{E}_k(\chi_k^\delta - \chi_\omega^\delta) + v\chi_\omega^\delta(\tilde{E}_k - \tilde{E}) \, d\boldsymbol{x}$$
$$- \int_\Omega \beta(\chi_k^\delta - \chi_\omega^\delta) \nabla E_b \cdot \nabla v + (\chi_k^\delta - \chi_\omega^\delta) E_b v \, d\boldsymbol{x},$$

and thus

$$|B_k[E_{b_k} - E_b, v]| \leq \|v\|_{H^1(\Omega)} \left(\left\| \tilde{E}_k - \tilde{E} \right\|_{H^1(\Omega)} \left(\delta + \left\| \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \right) \right.$$
$$\left. + \left\| \tilde{E}_k \right\|_{H^1(\Omega)} \left\| \chi_k^\delta - \chi_\omega^\delta \right\|_{L^\infty(\Omega)} + (\beta + 1) \|E_b\|_{H^1(\Omega)} \left\| \chi_k^\delta - \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \right).$$

By assumption $\tilde{E}_k \underset{H^1(\Omega)}{\longrightarrow} \tilde{E}$. Thus for every $\varepsilon > 0$ we can pick $K \in \mathbb{N}$ such that for $k \geq K$ we have $\left\| \tilde{E}_k \right\|_{H^1(\Omega)} < \left\| \tilde{E} \right\|_{H^1(\Omega)} + \varepsilon$. Hence we obtain

$$
\begin{aligned}
(4.80) \qquad |B_k[E_{\mathrm{b}_k} - E_\mathrm{b}, v]| \leq &\|v\|_{H^1(\Omega)} \left[\left\| \tilde{E}_k - \tilde{E} \right\|_{H^1(\Omega)} \left(\delta + \left\| \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \right) \right. \\
&+ \left(\left\| \tilde{E} \right\|_{H^1(\Omega)} + \varepsilon \right) \left\| \chi_k^\delta - \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \\
&\left. + (\beta + 1) \|E_\mathrm{b}\|_{H^1(\Omega)} \left\| \chi_k^\delta - \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \right].
\end{aligned}
$$

Let $\tau \left[\left\| \tilde{E}_k - \tilde{E} \right\|_{H^1(\Omega)}, \left\| \chi_k^\delta - \chi^\delta \right\|_{L^\infty(\Omega)} \right]$ denote the bracketed expression on the right hand side of (4.80). With $p = 1$, $r = \infty$ and $q = \infty$ we conclude by continuity of R_δ from Lemma 2 that $\|\chi_k - \chi_\omega\|_{L^1(\Omega)} \overset{k \to \infty}{\longrightarrow} 0$ implies $\left\| \chi_k^\delta - \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \overset{k \to \infty}{\longrightarrow} 0$. Thus

$$
(4.81) \qquad \tau \left[\left\| \tilde{E}_k - \tilde{E} \right\|_{H^1(\Omega)}, \left\| \chi_k^\delta - \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \right] \overset{k \to \infty}{\longrightarrow} 0.
$$

Observe further that due to $\chi_k^\delta \geq 0$, estimate (4.44) shows that all B_k are uniformly coercive with coercivity constant $C := \delta \min\{1, \beta\}$. Thus by setting $v = E_{\mathrm{b}_k} - E_\mathrm{b}$ in (4.80) and using (4.81) we get

$$
\|E_{\mathrm{b}_k} - E_\mathrm{b}\|_{H^1(\Omega)} \leq \frac{1}{C} \tau \left[\left\| \tilde{E}_k - \tilde{E} \right\|_{H^1(\Omega)}, \left\| \chi_k^\delta - \chi_\omega^\delta \right\|_{L^\infty(\Omega)} \right] \overset{k \to \infty}{\longrightarrow} 0,
$$

thus $E_{\mathrm{b}_k} \underset{H^1(\Omega)}{\longrightarrow} E_\mathrm{b}$. It follows analogously that $E_{\mathrm{f}_k} \underset{H^1(\Omega)}{\longrightarrow} E_\mathrm{f}$. Now define

$$
\begin{aligned}
h_k &:= \vartheta \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta}, & h &:= \vartheta \sqrt{\left| E_\mathrm{b} - \tilde{E} \right|^2 + \delta}, \\
g_k &:= \sqrt{\left| E_{\mathrm{f}_k} - \tilde{E}_k \right|^2 + \delta}, & g &:= \sqrt{\left| E_\mathrm{f} - \tilde{E} \right|^2 + \delta}.
\end{aligned}
$$

Mimicking the argumentation given in the beginning of the proof of Lemma 3 we see that h_k, h, g_k and g are $H^1(\Omega)$-functions. Furthermore, by using the identity (4.23) with

$\alpha = \left| E_{\mathrm{b}} - \tilde{E} \right|^2$ and $\alpha_k = \left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2$ we compute

$$\|h - h_k\|_{L^1(\Omega)} \leq \int_\Omega \vartheta \left| \sqrt{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 + \delta} - \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta} \right| d\boldsymbol{x}$$

$$= \int_\Omega \vartheta \left| \frac{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 - \left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2}{\sqrt{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta}} \right| d\boldsymbol{x}$$

(4.82)
$$= \int_\Omega \vartheta \left| \frac{(E_{\mathrm{b}} - E_{\mathrm{b}_k})(E_{\mathrm{b}_k} + E_{\mathrm{b}} - 2\tilde{E}_k) + (\tilde{E} - \tilde{E}_k)(\tilde{E} + \tilde{E}_k - 2E_{\mathrm{b}})}{\sqrt{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta}} \right| d\boldsymbol{x}$$

$$\leq \vartheta \left\| E_{\mathrm{b}} - E_{\mathrm{b}_k} \right\|_{L^2(\Omega)} \left\| \frac{E_{\mathrm{b}_k} + E_{\mathrm{b}} - 2\tilde{E}_k}{\sqrt{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta}} \right\|_{L^2(\Omega)} +$$

$$\vartheta \left\| \tilde{E} - \tilde{E}_k \right\|_{L^2(\Omega)} \left\| \frac{\tilde{E} + \tilde{E}_k - 2E_{\mathrm{b}}}{\sqrt{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta}} \right\|_{L^2(\Omega)} .$$

Since $\tilde{E}_k \xrightarrow[H^1(\Omega)]{} \tilde{E}$ by assumption and $E_{\mathrm{b}_k} \xrightarrow[H^1(\Omega)]{} E_{\mathrm{b}}$ as shown before we can pick $K \in \mathbb{N}$ such that for every $\varepsilon > 0$ and $k \geq K$ we have $\|E_{\mathrm{b}_k}\|_{L^2(\Omega)} < \|E_{\mathrm{b}}\|_{L^2(\Omega)} + \varepsilon$ and $\left\| \tilde{E}_k \right\|_{L^2(\Omega)} < \left\| \tilde{E} \right\|_{L^2(\Omega)} + \varepsilon$. Hence we obtain

(4.83)
$$\left\| \frac{E_{\mathrm{b}_k} + E_{\mathrm{b}} - 2\tilde{E}_k}{\sqrt{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta}} \right\|_{L^2(\Omega)} \leq \frac{1}{\sqrt{\delta}} \left(\|E_{\mathrm{b}}\|_{L^2(\Omega)} + \left\| \tilde{E} \right\|_{L^2(\Omega)} + \frac{3\varepsilon}{2} \right),$$

and

(4.84)
$$\left\| \frac{\tilde{E} + \tilde{E}_k - 2E_{\mathrm{b}}}{\sqrt{\left| E_{\mathrm{b}} - \tilde{E} \right|^2 + \delta} + \sqrt{\left| E_{\mathrm{b}_k} - \tilde{E}_k \right|^2 + \delta}} \right\|_{L^2(\Omega)} \leq \frac{1}{\sqrt{\delta}} \left(\|E_{\mathrm{b}}\|_{L^2(\Omega)} + \left\| \tilde{E} \right\|_{L^2(\Omega)} + \frac{\varepsilon}{2} \right).$$

Thus by combining (4.82) with (4.83) and (4.84) and using Young's inequality for convolutions (Appendix B.1) we get

$$\|(h - h_k) * \varphi_\delta\|_{L^1(\Omega)} \leq \|h - h_k\|_{L^1(\Omega)} \|\varphi_\delta\|_{L^2(\Omega)}$$

$$\leq \frac{\vartheta}{\sqrt{\delta}} \|E_{\mathrm{b}} - E_{\mathrm{b}_k}\|_{L^2(\Omega)} \left(\|E_{\mathrm{b}}\|_{L^2(\Omega)} + \left\| \tilde{E} \right\|_{L^2(\Omega)} + \frac{3\varepsilon}{2} \right) +$$

$$\frac{\vartheta}{\sqrt{\delta}} \left\| \tilde{E} - \tilde{E}_k \right\|_{L^2(\Omega)} \left(\|E_{\mathrm{b}}\|_{L^2(\Omega)} + \left\| \tilde{E} \right\|_{L^2(\Omega)} + \frac{\varepsilon}{2} \right),$$

hence $\quad \|(h - h_k) * \varphi_\delta\|_{L^1(\Omega)} \overset{k \to \infty}{\longrightarrow} 0.$ Similar considerations show that $\|(g - g_k) * \varphi_\delta\|_{L^1(\Omega)} \overset{k \to \infty}{\longrightarrow} 0$ as well since $E_{f_k} \underset{H^1(\Omega)}{\longrightarrow} E_f$. Now the claim is a consequence of the proof of Lemma 3. $\qquad \square$

Before we prove the first key result of this section we introduce the following notation. Recall that Φ, the mapping from Theorem 2, was used to extend the functional representation T of the Inner Iteration. Also Φ was shown to be continuous only with respect to some ξ (which was implicitly assumed to be a mollified edge map). Just like we extended T to bring in a dependence on $\tilde{E}$ by introducing the map $\mathcal{T}$ in Corollary 3 above we now adapt Φ to make the dependence on $\tilde{E}$ explicit. Thus let

$$(4.85) \qquad \Psi : \begin{cases} L^2(\Omega) \times H^1(\Omega) & \to L^2(\Omega), \\ (\xi, \tilde{E}) & \mapsto (M \circ \mathcal{T}) \left[\mathcal{H}[\xi], \tilde{E} \right], \end{cases}$$

where M and $\mathcal{H}$ are the mappings defined in Theorem 2. Note that since M maps sets onto their characteristic functions, $\Psi[\xi, \tilde{E}]$ is a characteristic function on Ω. For $N \in \mathbb{N}$ let

$$(4.86) \qquad \Psi_N[\xi, \tilde{E}] := \Psi \left[\Psi_{N-1}[\xi, \tilde{E}], \tilde{E} \right],$$

where $\Psi_0 := \mathrm{id}$. Note that (4.86) is the functional representation of the N-th step of the Inner Iteration as given by (4.85). Now we can combine the Inner Iteration (4.85) and the Outer Iteration (4.78) in the following manner. Let χ be a characteristic function on Ω. Then define

$$(4.87) \qquad \Xi : \begin{cases} L^2(\Omega) & \to L^2(\Omega), \\ \chi & \mapsto \Psi \left[\chi, \mathcal{I}[I[\chi^\delta]] \right], \end{cases}$$

and analogously for $N \in \mathbb{N}$ let

$$(4.88) \qquad \Xi_N[\chi] := \Psi_N \left[\chi, \mathcal{I}[I[\chi^\delta]] \right].$$

Thus the image of some initial edge map χ under Ξ_N is a characteristic function representing the N-fold evaluation of the Inner Iteration (4.85) using the fuzzy edge map (4.76) depending on χ. Now we can show the following.

Theorem 4. *Given Assumption 1 let χ be some characteristic function on Ω. Then the result of the inner iteration (4.85) after a finite number of steps depends continuously on the initial guess for χ, i.e., Ξ_N as given by (4.88) is continuous for every $N \in \mathbb{N}$.*

Proof. Let $\{\chi_k\}_{k \geq 1}$ be a sequence of characteristic functions converging $L^2(\Omega)$-strong to χ and let χ_k^δ and χ^δ denote the respective analytic mollifications in the sense of (4.38). Since

$$(4.89) \qquad \|\chi_k - \chi\|_{L^1(\Omega)} = \int_\Omega |\chi_k - \chi| \, d\boldsymbol{x} = \int_\Omega |\chi_k - \chi|^2 \, d\boldsymbol{x} = \|\chi_k - \chi\|_{L^2(\Omega)}^2 \overset{k \to \infty}{\longrightarrow} 0,$$

we have $\chi_k \underset{L^1(\Omega)}{\longrightarrow} \chi$ and thus by Lemma 5, $I[\chi_k^\delta] \underset{H^2(\Omega)}{\longrightarrow} I[\chi^\delta]$. For the sake of brevity let $I_k := [\chi_k^\delta]$ and $I_\chi := I[\chi^\delta]$. According to Lemma 6 the mapping $\mathcal{I}$ is continuous thus

---

Algorithm 4.1 Adaptation of Algorithm 3.1

> **Input:** $\tilde{I} \in L^2(\Omega)$, a characteristic function χ, $n, N \in \mathbb{N}$, $\beta, \vartheta > 0$ and $0 < \delta \ll 1$
> **Output:** χ

1: **for** $i = 1, \ldots, n$ **do**
2:     **Evaluate** $\tilde{E} = \mathcal{I}[I[R_\delta[\chi]]]$
3:     **Set** $\chi = \Xi_N[\chi]$ for $N \in \mathbb{N}$
4: **end for**

$\mathcal{I}[I_k] \underset{H^1(\Omega)}{\longrightarrow} \mathcal{I}[I_\chi]$. Hence with $\tilde{E}_k := \mathcal{I}[I_k]$ and $\tilde{E} := \mathcal{I}[I_\chi]$ we have $\tilde{E}_k \underset{H^1(\Omega)}{\longrightarrow} \tilde{E}$. Since by assumption $\chi_k \underset{L^2(\Omega)}{\longrightarrow} \chi$ we conclude by continuity of $\mathcal{H}$ (compare the proof of Theorem 2) that $|\omega \bigtriangleup \omega_k| \overset{k\to\infty}{\longrightarrow} 0$ with $\omega := \mathcal{H}[\chi]$ and $\omega_k := \mathcal{H}[\chi_k]$. Thus Corollary 3 ensures that $\left|\mathcal{T}[\omega, \tilde{E}] \bigtriangleup \mathcal{T}[\omega_k, \tilde{E}_k]\right| \overset{k\to\infty}{\longrightarrow} 0$ since $\mathcal{T}$ is continuous. Again by the proof of Theorem 2 it follows that M is continuous and hence Ψ is continuous as well. Since compositions of continuous maps are also continuous we conclude inductively that Φ_N is continuous. Thus by the argumentation given above, Ξ_N is continuous for all $N \in \mathbb{N}$. $\qquad\square$

With the tools introduced in this section, particularly (4.78) and (4.88) we can formulate a theoretical version of Algorithm 3.1. As discussed above the function evaluations (4.78) and (4.88) approximate Outer and Inner Iteration as seen in Algorithm 3.1. Thus Algorithm 4.1 can be seen as an approximation to Algorithm 3.1 for vanishingly small values of δ. Finally we state a stability result for Algorithm 4.1 that is an immediate consequence of Theorem 4.

Corollary 4. *The result of Algorithm 4.1 depends continuously on the initial binary edge map χ.*

Proof. Let $\{\chi_k\}_{k\geq 1}$ be a sequence of characteristic functions converging $L^2(\Omega)$-strong to χ. By Theorem 4, Ξ_N is continuous; thus, $\Xi_N[\chi_k] \underset{L^2(\Omega)}{\longrightarrow} \Xi_N[\chi]$. Since compositions of continuous maps are continuous we conclude that $\Xi_N[\Xi_N[\chi_k]] \underset{L^2(\Omega)}{\longrightarrow} \Xi_N[\Xi_N[\chi]]$ and inductively $\Xi_N^n[\chi_k] \underset{L^2(\Omega)}{\longrightarrow} \Xi_N^n[\chi]$ where we used the notation $\Xi^n := \Xi \circ \Xi^{n-1}$ with $\Xi^0 := \mathrm{id}$.
$\qquad\square$

This ends the analysis of an infinite dimensional formulation of Algorithm 3.1. The next chapter addresses the practical implementation of Algorithm 3.1 using finite elements.

<div style="text-align: right; font-size: 4em; font-weight: bold;">5</div>

Numerical Approximation

This chapter is devoted to the development of a consistent discretization of Algorithm 3.1. We will rely on the *method of finite elements* to achieve this objective. As detailed in Section 2.1 we only consider digital grayscale images. For the sake of simplicity we restrict ourselves further to quadratic images of resolution $N \times N$.

We start by defining spline approximation spaces in 1D and use tensor products to construct the respective spaces in two dimensions. Thus consider the interval $\Omega^{1\mathrm{D}} := (0, 1)$ and let $\Omega_i^{1\mathrm{D}} := (x_{i-1}, x_i)$ be a grid on $\Omega^{1\mathrm{D}}$ with nodes $x_i = ih$ and stepsize $h = 1/N$ for $i = 0, \ldots, N$. We introduce the *spline bases*

$$S_h^{(m)}(\Omega^{1\mathrm{D}}) := \left\{ s \in \mathcal{P}^m([x_{i-1}, x_i]) \,|\, s \in C^{m-1}(\Omega^{1\mathrm{D}}), \, i = 1, \ldots N \right\}, \quad m = 0, 1, \ldots,$$

where $\mathcal{P}^m([x_{i-1}, x_i])$ denotes the space of polynomials of degree m on the interval $[x_{i-1}, x_i]$. For $m = 0$ we mean cellwise constant functions. Let π_m denote the *canonical spline* of order m, that is

$$(5.1) \qquad\qquad \pi_m(x) = (\pi_{m-1} * \pi_0)(x), \quad m = 1, 2, \ldots,$$

with π_0 being the characteristic function of the interval $[0, 1]$; see Figure 5.1 and compare, e.g., [DH02, Sec. 7.4.1]. We use

$$\hat{s}_{i+m+1}^{(m)}(x) := \pi \left(\frac{x - x_i}{h} \right), \quad i = -m, \ldots, N - 1,$$

as basis functions for $S_h^{(m)}(\Omega^{1\mathrm{D}})$. For $\Omega = (0, 1)^2$ we interpret $S_h^{(m)}(\Omega)$ as tensor products of these spline bases. For an exhaustive review of the construction of approximation spaces in various dimensions see [GR05, Chap. 4].

Remark 2. *Let $m \in \mathbb{N}_0$ and identify $H^0(\Omega)$ with $L^2(\Omega)$. Then the spaces $S_h^{(m)}(\Omega)$ are dense in $H^m(\Omega)$ [Sch81]. Further, $\{s_i^{(m)}\}_{i=1}^{(N+m)^2}$ is a basis of $S_h^{(m)}(\Omega)$.*

Having introduced spaces and notation we may now start our considerations by looking at the raw image $\tilde{I}$. Since we work with digital grayscale images of size $N \times N$ we only have access to N^2 pixelwise averaged intensity values. Thus let $\tilde{I}_h$ denote the spline representation of $\tilde{I}$. In the just introduced terminology the pixelwise averaged approximation $\tilde{I}_h$ is thus a member of $S_h^{(0)}(\Omega)$. Hence let $\{\tilde{I}_{h,i}\}_{i=1}^{N^2}$ denote the intensity values of $\tilde{I}_h$ in
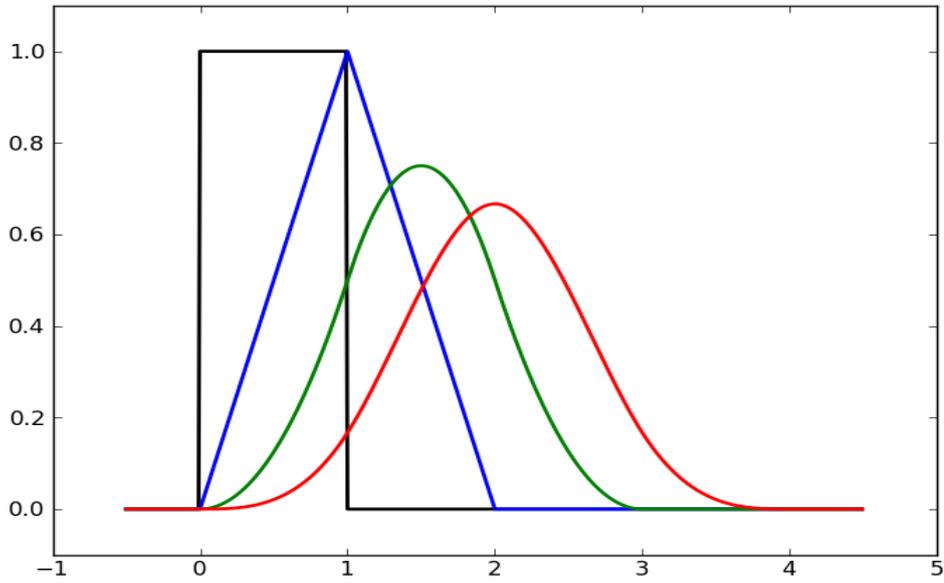
Figure 5.1: The canonical B-splines. Shown is the piecewise constant ($m = 0$, black), linear ($m = 1$, blue), quadratic ($m = 2$, green) and cubic ($m = 3$, red) B-spline as given by (5.1).

lexicographic ordering. Then

$$(5.2) \qquad \tilde{I}_h = \sum_{i=1}^{N^2} \tilde{I}_{h,i} s_i^{(0)}.$$

Similarly, let $\chi_h \in S_h^{(0)}(\Omega)$ given by

$$(5.3) \qquad \chi_h = \sum_{i=1}^{N^2} \chi_{h,i} s_i^{(0)}.$$

be the spline representation of some characteristic function χ. We make the following assumption governing the approximation quality of these discretizations.

Assumption 5. *The pixelwise data approximation $\tilde{I}_h$ as well as the representation χ_h of some edge map χ satisfy*

$$\left\| \tilde{I}_h - \tilde{I} \right\|_{L^\infty(\Omega)} \xrightarrow{h \to 0} 0, \quad \left\| \chi_h - \chi \right\|_{L^\infty(\Omega)} \xrightarrow{h \to 0} 0.$$

Recall that Algorithm 3.1 involves the solution of various minimization problems. Thus in practice we are interested in (approximately) solving the optimality conditions (4.3), (4.8) and (4.63). Hence in the following we show well posedness and convergence of respective finite element formulations.

5.1 Analysis of a Finite Element Approach for the Outer Iteration

We start by considering the Outer Iteration and in particular the weak formulation of the necessary optimality condition associated to the minimization problem (4.62). Note that $I_{\mathrm{s}} \in H^2(\Omega)$ and thus following Remark 2 we choose $S_h^{(2)}(\Omega)$ as approximation space. Hence a finite element discretization of (4.63) is to compute $I_{\mathrm{s},h} \in S_h^{(2)}(\Omega)$ so that

$$(5.4) \qquad A_h[I_{\mathrm{s},h}, v_h] = d_h[v_h], \quad \forall v_h \in S_h^{(2)}(\Omega),$$

where $A_h : H^2(\Omega) \times H^2(\Omega) \to \mathbb{R}$ is defined by

$$(5.5) \qquad A_h[u,v] := \int_\Omega \beta(\delta + \chi_h)\nabla^2 u : \nabla^2 v + (\delta + \chi_h)uv \, d\boldsymbol{x},$$

and $d_h : H^2(\Omega) \to \mathbb{R}$ is given by

$$(5.6) \qquad d_h[v] := \int_\Omega (\delta + \chi_h)\tilde{I}_h v \, d\boldsymbol{x}.$$

We show that $I_{\mathrm{s},h}$ is indeed the unique solution to (5.4).

Theorem 5. *Given Assumptions 1 and 5 there exists a unique $I_{s,h} \in S_h^{(2)}(\Omega)$ satisfying* (5.4).

Proof. The claim is an immediate consequence of the proof of Theorem 3. As $\|\chi_h\|_{L^\infty(\Omega)} \leq 1$ we see that d_h is bounded

$$d_h[v] \leq (\delta + 1)\left\|\tilde{I}_h\right\|_{L^\infty(\Omega)} \|v\|_{H^2(\Omega)} \sqrt{|\Omega|}.$$

Similarly A_h is bounded as well

$$|A_h[u,v]| \leq (\delta + 1)(\beta + 1) \|u\|_{H^2(\Omega)} \|v\|_{H^2(\Omega)}.$$

Since $\chi_h \geq 0$ we can establish estimate (4.66) for A_h as well. Then following the argumentation in the proof of Theorem 3, A_h is also coercive on $H^2(\Omega)$ for every $h > 0$. Since by Remark 2, $S_h^{(2)}(\Omega) \subset H^2(\Omega)$, the Lax–Milgram Lemma (Appendix B.2) shows unique solvability of (5.4) which proves the claim. $\qquad\square$

Having shown well posedness of the problem defining $I_{\mathrm{s},h}$ we can now prove convergence of the discretization.

Theorem 6. *Given Assumptions 1 and 5 let $I_{s,h}$ be the unique solution to (5.4) and let I_s be uniquely defined by the weak formulation (4.63). Then $I_{s,h} \xrightarrow[H^2(\Omega)]{} I_s$ as $h \to 0$.*

Proof. Observe that for any $v \in H^2(\Omega)$ we have

$$A_h[I_{\mathrm{s},h} - I_{\mathrm{s}}, v] = \int_\Omega \delta v(\tilde{I}_h - \tilde{I}) + v\tilde{I}_h(\chi_h - \chi) + v\chi(\tilde{I}_h - \tilde{I}) \, d\boldsymbol{x}$$
$$- \int_\Omega \beta(\chi_h - \chi)\nabla^2 I_{\mathrm{s}} : \nabla^2 v + (\chi_h - \chi)I_{\mathrm{s}}v \, d\boldsymbol{x},$$

and thus

$$|A_h[I_{s,h} - I_s, v]| \leq \|v\|_{H^2(\Omega)} \left(\left\|\tilde{I}_h - \tilde{I}\right\|_{L^\infty(\Omega)} \sqrt{|\Omega|} \left(\delta + \|\chi\|_{L^\infty(\Omega)}\right) \right.$$
$$+ \left\|\tilde{I}_h\right\|_{L^\infty(\Omega)} \|\chi_h - \chi\|_{L^\infty(\Omega)} \sqrt{|\Omega|}$$
$$\left. + (\beta + 1) \|I_s\|_{H^2(\Omega)} \|\chi_h - \chi\|_{L^\infty(\Omega)} \right).$$

By Assumption 5, $\tilde{I}_h \xrightarrow[L^\infty(\Omega)]{} \tilde{I}$, and hence for every $\varepsilon > 0$ we may pick $K \in \mathbb{N}$ such that for $k \geq K$ we have $\left\|\tilde{I}_h\right\|_{L^\infty(\Omega)} < \left\|\tilde{I}\right\|_{L^\infty(\Omega)} + \varepsilon$. Thus we obtain

(5.7)
$$|A_h[I_{s,h} - I_s, v]| \leq \|v\|_{H^2(\Omega)} \left[\left\|\tilde{I}_h - \tilde{I}\right\|_{L^\infty(\Omega)} \sqrt{|\Omega|} \left(\delta + \|\chi\|_{L^\infty(\Omega)}\right) \right.$$
$$+ \left(\left\|\tilde{I}\right\|_{L^\infty(\Omega)} + \varepsilon \right) \|\chi_h - \chi\|_{L^\infty(\Omega)} \sqrt{|\Omega|}$$
$$\left. + (\beta + 1) \|I_s\|_{H^2(\Omega)} \|\chi_h - \chi\|_{L^\infty(\Omega)} \right].$$

Let $\tau \left[\left\|\tilde{I}_h - \tilde{I}\right\|_{L^\infty(\Omega)}, \|\chi_h - \chi\|_{L^\infty(\Omega)} \right]$ denote the bracketed expression on the right hand side of (5.7). According to Assumption 5, $\tilde{I}_h \xrightarrow[L^\infty(\Omega)]{} \tilde{I}$ and $\chi_h \xrightarrow[L^\infty(\Omega)]{} \chi$, thus,

(5.8)
$$\tau \left[\left\|\tilde{I}_h - \tilde{I}\right\|_{L^\infty(\Omega)}, \|\chi_h - \chi\|_{L^\infty(\Omega)} \right] \xrightarrow{h \to 0} 0.$$

Let further $c_s > 0$ denote again the coercivity constant of A_s. According to the proof of Theorem 5, A_h is also coercive on $H^2(\Omega)$ for every $h > 0$. Hence let similarly c_h denote the coercivity constant of A_h. Then

(5.9)
$$c_h = \inf_{v \in H^2(\Omega)} \frac{|A_h[v, v]|}{\|v\|_{H^2(\Omega)}^2} > 0, \quad \forall h > 0.$$

Since

$$|A_h[v, v] - A_s[v, v]| = \left| \int_\Omega (\chi_h - \chi) \left|\nabla^2 v\right|^2 + (\chi_h - \chi) v^2 d\boldsymbol{x} \right|$$
$$\leq \|\chi_h - \chi\|_{L^\infty(\Omega)} \|v\|_{H^2(\Omega)}^2 \xrightarrow{h \to 0} 0,$$

by adding and subtracting $A[v, v]$ in (5.9) we conclude that $c_h \xrightarrow{h \to 0} c_s$. Thus we can choose $\eta > 0$ such that $c_h > \eta > 0$ for all h small enough. Then for $v = I_s - I_{s,h}$ and h sufficiently small we obtain
$$\|I_s - I_{s,h}\|_{H^2(\Omega)}^2 \leq \frac{1}{\eta} A_h[I_s - I_{s,h}, I_s - I_{s,h}].$$

This together with (5.7) and (5.8) finally yields

$$\|I_{\mathrm{s}} - I_{\mathrm{s},h}\|_{H^2(\Omega)} \leq \frac{1}{\eta} A_h[I_{\mathrm{s}} - I_{\mathrm{s},h}, I_{\mathrm{s}} - I_{\mathrm{s},h}] \leq \frac{1}{\eta}\tau \left[\left\| \tilde{I}_h - \tilde{I} \right\|_{L^\infty(\Omega)}, \|\chi_h - \chi\|_{L^\infty(\Omega)} \right] \xrightarrow{h \to 0} 0,$$

which proves the claim. $\qquad \square$

In the next section we prove analogous results for the Inner Iteration.

5.2 Analysis of a Finite Element Approach for the Inner Iteration

Before we focus on a finite element formulation of the weak problems associated to the Inner Iteration we have to properly set up the link between the discrete Outer and the discrete Inner Iteration. In other words, we have to set up a discretization of the fuzzy edge map $\tilde{E}$. Thus following Algorithm 3.1 let $I_{\mathrm{s},h}$ be the unique solution to (5.4) (whose existence is guaranteed by Theorem 5) and define

$$(5.10) \qquad\qquad\qquad \tilde{E}_h := |\nabla I_{\mathrm{s},h}|.$$

Observe that for solving (4.3) and (4.8) we only required $\tilde{E} \in L^2(\Omega)$ (compare Theorem 1). Hence we define $\tilde{E}_h$ as above instead of following (4.76). The minimization problems (4.2) and (4.7) and thus their associated optimality conditions (4.3) and (4.8) respectively are posed on $H^1(\Omega)$. Thus by Remark 2 we choose $S_h^{(1)}(\Omega)$ as approximation space. Then a finite element discretization of (4.3) is to compute $E_{\mathrm{b},h} \in S_h^{(1)}(\Omega)$ so that

$$(5.11) \qquad\qquad B_{\mathrm{b},h}[E_{\mathrm{b},h}, v_h] = f_{\mathrm{b},h}[v_h], \quad \forall v_h \in S_h^{(1)}(\Omega),$$

where the bilinear form $B_{\mathrm{b},h} : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ is defined by

$$(5.12) \qquad\qquad B_{\mathrm{b},h}[u,v] := \int_\Omega \beta(\delta + \chi_h)\nabla u \cdot \nabla v + (\delta + \chi_h)uv \, d\boldsymbol{x},$$

and the linear functional $f_{\mathrm{b},h} : H^1(\Omega) \to \mathbb{R}$ is given by

$$f_{\mathrm{b},h}[v] := \int_\Omega (\delta + \chi_h)\tilde{E}_h v \, d\boldsymbol{x}.$$

Similarly (4.8) is discretized by computing $E_{\mathrm{f},h} \in H^1(\Omega)$ such that

$$(5.13) \qquad\qquad B_{\mathrm{f},h}[E_{\mathrm{f},h}, v_h] = f_{\mathrm{f},h}[v_h], \quad \forall v_h \in S_h^{(1)}(\Omega),$$

with $B_{\mathrm{f},h} : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ given by

$$(5.14) \qquad\qquad B_{\mathrm{f},h}[u,v] := \int_\Omega \beta(\delta + 1 - \chi_h)\nabla u \cdot \nabla v + (\delta + 1 - \chi_h)uv \, d\boldsymbol{x},$$

and $f_{\mathrm{f},h} : H^1(\Omega) \to \mathbb{R}$ defined by

$$f_{\mathrm{f},h}[v] := \int_\Omega (\delta + 1 - \chi_h)\tilde{E}_h v \, d\boldsymbol{x}.$$

The following result shows that $E_{b,h}$ and $E_{f,h}$ are the unique solutions to (5.11) and (5.13) respectively.

Theorem 7. *Given Assumptions 1 and 5, $E_{b,h}$ and $E_{f,h}$ are uniquely determined by (5.11) and (5.13) respectively.*

Proof. Since both $\|\chi_h\|_{L^\infty(\Omega)}, \|1 - \chi_h\|_{L^\infty(\Omega)} \leq 1$ and $\chi_h, 1 - \chi_h \geq 0$ we only show existence and uniqueness of $E_{b,h}$. Unique solvability of (5.13) follows from replacing χ_h by $1 - \chi_h$ in the arguments given below.

The claim follows immediately from Theorem 1. Boundedness of $f_{b,h}$ is readily established

$$f_{b,h}[v] \leq (\delta + 1) \left\| \tilde{E}_h \right\|_{L^2(\Omega)} \|v\|_{H^1(\Omega)},$$

similarly we see that $B_{b,h}$ is also bounded

$$|B_{b,h}[u, v]| \leq (\delta + 1)(\beta + 1) \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}.$$

Estimate (4.11) shows that $B_{b,h}$ is also coercive for every $h > 0$. By Remark 2 $S_h^{(1)}(\Omega) \subset H^1(\Omega)$; thus, we conclude by the Lax–Milgram Lemma that $E_{b,h}$ is uniquely defined by (5.11). □

Before we can show convergence of the discretizations $E_{b,h}$ and $E_{f,h}$ we have to assure that the discrete fuzzy edge maps (5.10) converges.

Corollary 6. *Given Assumption 1 consider $\tilde{E}$ used in Algorithm 3.1 and $\tilde{E}_h$ given by (5.10). Then $\tilde{E}_h \underset{L^2(\Omega)}{\longrightarrow} \tilde{E}$.*

Proof. We compute

$$\begin{aligned}
\left\| \tilde{E} - \tilde{E}_h \right\|_{L^2(\Omega)} &= \| |\nabla I_s| - |\nabla I_{s,h}| \|_{L^2(\Omega)} \\
&\leq \|\nabla I_s - \nabla I_{s,h}\|_{L^2(\Omega)} \\
&\leq \|I_s - I_{s,h}\|_{H^1(\Omega)} \\
&\leq \|I_s - I_{s,h}\|_{H^2(\Omega)} \overset{h \to 0}{\longrightarrow} 0,
\end{aligned}$$

by Theorem 6. □

Now we can prove convergence of $E_{b,h}$ and $E_{f,h}$ to E_b and E_f respectively.

Theorem 8. *Given Assumptions 1 and 5 let $E_{b,h}$ be the unique solution to (5.11) and let E_b be uniquely defined by (4.3). Similarly let $E_{f,h}$ and E_f be uniquely determined by (5.13) and (4.8) respectively. Then $E_{b,h} \underset{H^1(\Omega)}{\longrightarrow} E_b$ and $E_{f,h} \underset{H^1(\Omega)}{\longrightarrow} E_f$ as $h \to 0$.*

Proof. Again, we only prove convergence of $E_{b,h}$ to E_b. Replacing χ_h by $1 - \chi_h$ in the arguments given below show $E_{f,h} \underset{H^1(\Omega)}{\longrightarrow} E_f$.

For $v \in H^1(\Omega)$ we have

$$B_{\mathrm{b},h}[E_{\mathrm{b},h} - E_{\mathrm{b}}, v] = \int_\Omega \delta v \left(\tilde{E}_h - \tilde{E} \right) + v \tilde{E}_h \left(\chi_h - \chi \right) + v \chi \left(\tilde{E}_h - \tilde{E} \right) \, d\boldsymbol{x}$$
$$- \int_\Omega \beta \left(\chi_h - \chi \right) \nabla E_{\mathrm{b}} \nabla v + \left(\chi_h - \chi \right) E_{\mathrm{b}} v \, d\boldsymbol{x},$$

and thus

$$|B_{\mathrm{b},h}[E_{\mathrm{b},h} - E_{\mathrm{b}}, v]| \leq \|v\|_{H^1(\Omega)} \left(\left\| \tilde{E}_h - \tilde{E} \right\|_{L^2(\Omega)} \left(\delta + \|\chi\|_{L^\infty(\Omega)} \right) \right.$$
$$+ \left\| \tilde{E}_h \right\|_{L^2(\Omega)} \|\chi_h - \chi\|_{L^\infty(\Omega)}$$
$$\left. + (\beta + 1) \|E_{\mathrm{b}}\|_{H^1(\Omega)} \|\chi_h - \chi\|_{L^\infty(\Omega)} \right).$$

According to Corollary 6 $\tilde{E}_h \xrightarrow[L^2(\Omega)]{} \tilde{E}$ thus for any $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that for all $k \geq K$ we have $\left\| \tilde{E}_h \right\|_{L^2(\Omega)} < \left\| \tilde{E} \right\|_{L^2(\Omega)} + \varepsilon$. Hence we get

$$|B_{\mathrm{b},h}[E_{\mathrm{b},h} - E_{\mathrm{b}}, v]| \leq \|v\|_{H^1(\Omega)} \left[\left\| \tilde{E}_h - \tilde{E} \right\|_{L^2(\Omega)} \left(\delta + \|\chi\|_{L^\infty(\Omega)} \right) \right.$$

(5.15)
$$+ \left(\left\| \tilde{E} \right\|_{L^2(\Omega)} + \varepsilon \right) \|\chi_h - \chi\|_{L^\infty(\Omega)}$$
$$\left. + (\beta + 1) \|E_{\mathrm{b}}\|_{H^1(\Omega)} \|\chi_h - \chi\|_{L^\infty(\Omega)} \right].$$

Let $\tau \left[\left\| \tilde{E}_h - \tilde{E} \right\|_{L^2(\Omega)}, \|\chi_h - \chi\|_{L^\infty(\Omega)} \right]$ denote the bracketed expression on the right hand side of (5.15). By Assumption 5 and Corollary 6 we have $\chi_h \xrightarrow[L^\infty(\Omega)]{} \chi$ and $\tilde{E}_h \xrightarrow[L^2(\Omega)]{} \tilde{E}$ respectively. Hence

(5.16)
$$\tau \left[\left\| \tilde{E}_h - \tilde{E} \right\|_{L^2(\Omega)}, \|\chi_h - \chi\|_{L^\infty(\Omega)} \right] \xrightarrow{h \to 0} 0.$$

Since $\chi_h \geq 0$ estimate (4.11) shows that $B_{\mathrm{b},h}$ is uniformly coercive for all $h > 0$ with the coercivity constant $C := \delta \min\{\beta, 1\}$. Hence for $v = E_{\mathrm{b},h} - E_{\mathrm{b}}$ relations (5.15) and (5.16) finally give

$$\|E_{\mathrm{b},h} - E_{\mathrm{b}}\|_{H^1(\Omega)} \leq \frac{1}{C} B_{\mathrm{b},h}[E_{\mathrm{b},h} - E_{\mathrm{b}}, E_{\mathrm{b},h} - E_{\mathrm{b}}]$$
$$\leq \frac{1}{C} \tau \left[\left\| \tilde{E}_h - \tilde{E} \right\|_{L^2(\Omega)}, \|\chi_h - \chi\|_{L^\infty(\Omega)} \right] \xrightarrow{h \to 0} 0,$$

which proves the claim. $\qquad \square$

Now we are ready to establish a discrete formulation of Algorithm 3.1.

5.3 A Discrete Formulation of Algorithm 3.1

Having shown solvability and convergence of a finite element approach for the Inner as well as for the Outer Iteration we can now give a more detailed and practical pseudo-code for Algorithm 3.1. Recall that we assumed $\tilde{I}_h$ to be the finite element representation of a quadratic $N \times N$ grayscale image. Thus $h = 1/N$ in the x- and y- directions since $\Omega = (0,1)^2$. Algorithm 3.1 was designed to compute a binary edge map in an iterative manner. Thus it obviously requires the edge map to be updated regularly. In view of computational efficiency we thus seek to decouple terms in (5.4), (5.11) and (5.13) that are not affected by χ_h from the rest. Hence for $u_h, v_h \in S_h^{(m)}(\Omega)$ with $m = 1, 2$ let

$$c_{\chi_h}[u_h, v_h] := \int_\Omega \chi_h u_h v_h \, d\boldsymbol{x}.$$

We start by rewriting (5.4) as a linear equation system. Thus let $\{s_i^{(2)}\}_{i=1}^{(N+2)^2}$ be the basis of $S_h^{(2)}(\Omega)$ addressed in Remark 2. Observe that solving (5.4) means computing coefficients $I_{\mathrm{s},h,i}$ such that

$$I_{\mathrm{s},h} = \sum_{i=1}^{(N+2)^2} I_{\mathrm{s},h,i} s_i^{(2)}.$$

Thus let $\boldsymbol{I}_{\mathrm{s}} := \{I_{\mathrm{s},h,i}\}_{i=1}^{(N+2)^2} \in \mathbb{R}^{(N+2)^2}$ denote the vector of coefficient values. Recall the form A_h given by (5.5). Following the idea of separating terms depending on χ_h from the rest we define

$$a_{\chi_h}[u_h, v_h] := \int_\Omega \chi_h \nabla^2 u_h : \nabla^2 v_h \, d\boldsymbol{x}.$$

Consequently we introduce the following notation for the bending matrix associated to A_h

$$\boldsymbol{A}[\chi_h] := \left\{ a_{\chi_h}[s_i^{(2)}, s_j^{(2)}] \right\}_{i,j=1}^{(N+2)^2} \in \mathbb{R}^{(N+2)^2 \times (N+2)^2}.$$

which reflects its dependence on χ_h. Similarly the Gram matrix for A_h is denoted by

$$\boldsymbol{G}[\chi_h] := \left\{ c_{\chi_h}[s_i^{(2)}, s_j^{(2)}] \right\}_{i,j=1}^{(N+2)^2} \in \mathbb{R}^{(N+2)^2 \times (N+2)^2}.$$

Stencils illustrating the structure of both matrices $\boldsymbol{A}[\chi_h]$ and $\boldsymbol{G}[\chi_h]$ are given in Appendix C.2. Let further $\tilde{\boldsymbol{I}} := \{\tilde{I}_{h,i}\}_{i=1}^{N^2}$ denote the vector of intensity values of $\tilde{I}_h$ as given by (5.2). In contrast to the bending and Gram matrix the right hand side d_h of (5.6) has a simple representation terms of a projection of $\tilde{\boldsymbol{I}}$. Thus set $v = s_k^{(2)}$ in d_h and note that $s_i^{(0)} s_j^{(0)} = \mathfrak{d}_{i,j}$ (where $\mathfrak{d}_{i,j}$ denotes the Kronecker delta). Then by using the spline representations of $\tilde{I}_h$ and χ_h we compute

$$(5.17) \quad d_h[s_k^{(2)}] = \sum_{i,j=1}^{N^2} \int_\Omega (\delta + \chi_{h,i}) \, s_i^{(0)} \tilde{I}_{h,j} s_j^{(0)} s_k^{(2)} \, d\boldsymbol{x} = \sum_{i=1}^{N^2} \int_\Omega (\delta + \chi_{h,i}) \, \tilde{I}_{h,i} s_i^{(0)} s_k^{(2)} \, d\boldsymbol{x}$$

We introduce the projection matrix that maps quadratic splines to constants

$$\boldsymbol{P}_2 := \left\{ \int_\Omega s_i^{(0)} s_j^{(2)} \right\}_{\substack{i=1,\dots,N^2 \\ j=1,\dots,(N+2)^2}} \in \mathbb{R}^{N^2 \times (N+2)^2}.$$

The stencil for $\boldsymbol{P}_2$ can be found in Appendix C.2 as well. Let further $\boldsymbol{\chi} := \{\chi_{h,i}\}_{i=1}^{N^2} \in \mathbb{R}^{N^2}$ denote the vector of coefficients of χ_h as given by (5.3). Then (5.17) is the k-th component of the vector

$$\boldsymbol{P}_2^\top \left((\delta + \boldsymbol{\mathcal{D}}[\boldsymbol{\chi}]) \tilde{\boldsymbol{I}} \right),$$

where $\boldsymbol{\mathcal{D}}[\boldsymbol{\chi}] \in \mathbb{R}^{N^2 \times N^2}$ denotes a diagonal matrix with the values $\chi_{h,i}$ of χ_h on its main diagonal. Thus we eventually obtain the following expression for (5.4)

$$(\beta \boldsymbol{A}[\chi_h] + \boldsymbol{G}[\chi_h] + \delta (\beta \boldsymbol{A}[1] + \boldsymbol{G}[1])) \, \boldsymbol{I}_{\mathrm{s}} = \boldsymbol{P}_2^\top \left((\delta + \boldsymbol{\mathcal{D}}[\boldsymbol{\chi}]) \tilde{\boldsymbol{I}} \right).$$

In a similar manner we can deduce linear equation systems for (5.11) and (5.13). Thus let $\{s_i^{(1)}\}_{i=1}^{(N+1)^2}$ denote the basis of $S_h^{(1)}(\Omega)$. We want to determine $\boldsymbol{E}_{\mathrm{b}} := \{E_{\mathrm{b},h,i}\}_{i=1}^{(N+1)^2} \in \mathbb{R}^{(N+1)^2}$ and $\boldsymbol{E}_{\mathrm{f}} := \{E_{\mathrm{f},h,i}\}_{i=1}^{(N+1)^2} \in \mathbb{R}^{(N+1)^2}$ such that

$$E_{\mathrm{b},h} = \sum_{i=1}^{(N+1)^2} E_{\mathrm{b},h,i} s_i^{(1)} \text{ and } E_{\mathrm{f},h} = \sum_{i=1}^{(N+1)^2} E_{\mathrm{f},h,i} s_i^{(1)}.$$

Following the strategy above we start by considering (5.11). We analogously introduce

$$b_{\chi_h}[u_h, v_h] := \int_\Omega \chi_h \nabla u_h \cdot \nabla v_h \, d\boldsymbol{x},$$

and thus obtain the stiffness matrix

$$\boldsymbol{B}[\chi_h] := \left\{ b_{\chi_h}[s_i^{(1)}, s_j^{(1)}] \right\}_{i,j=1}^{(N+1)^2} \in \mathbb{R}^{(N+1)^2 \times (N+1)^2},$$

for $B_{\mathrm{b},h}$ given by (5.12). The Gram matrix associated to $B_{\mathrm{b},h}$ is given by

$$\boldsymbol{C}[\chi_h] := \left\{ c_{\chi_h}[s_i^{(1)}, s_j^{(1)}] \right\}_{i,j=1}^{(N+1)^2} \in \mathbb{R}^{(N+1)^2 \times (N+1)^2}.$$

Again, the stencils structuring both matrices are given in Appendix C.2. Similar to the situation above we can express the right hand side $f_{\mathrm{b},h}$ of (5.11) in terms of $\tilde{E}_h$ and a projection. Note that since $\tilde{E}_h = |\nabla I_{\mathrm{s},h}|$ we have $\tilde{E}_h \in L^2(\Omega)$. Thus let $\mathfrak{P}_h$ denote the $L^2(\Omega)$-projection onto $S_h^{(1)}(\Omega)$, i.e., $\mathfrak{P}_h : L^2(\Omega) \to S_h^{(1)}(\Omega)$. Then by using the spline representations of $\chi_h \in S_h^{(0)}(\Omega)$ and $(\mathfrak{P}_h \tilde{E}_h) \in S_h^{(1)}(\Omega)$ with $\tilde{\boldsymbol{E}} := \{(\mathfrak{P}_h \tilde{E}_h)_i\}_{i=1}^{(N+1)^2} \in \mathbb{R}^{(N+1)^2}$ denoting the coefficient vector of $\mathfrak{P}_h \tilde{E}_h$ we compute

$$(5.18) \qquad f_{\mathrm{b},h}[s_k^{(1)}] = \sum_{j=1}^{(N+1)^2} \int_\Omega \sum_{i=1}^{N^2} (\delta + \chi_{h,i}) \, s_i^{(0)} (\mathfrak{P}_h \tilde{E}_h)_j s_j^{(1)} s_k^{(1)} \, d\boldsymbol{x}.$$

Thus (5.18) is the k-th component of the vector

$$(\delta \boldsymbol{C}[1] + \boldsymbol{C}[\chi_h]) \, \tilde{\boldsymbol{E}},$$

and thus we obtain the following linear system that is equivalent to (5.11)

$$(\beta \boldsymbol{B}[\chi_h] + \boldsymbol{C}[\chi_h] + \delta \, (\beta \boldsymbol{B}[1] + \boldsymbol{C}[1])) \, \boldsymbol{E}_{\mathrm{b},h} = (\delta \boldsymbol{C}[1] + \boldsymbol{C}[\chi_h]) \, \tilde{\boldsymbol{E}}.$$

Analogously (5.13) leads to the system

$$(\beta \boldsymbol{B}[1 - \chi_h] + \boldsymbol{C}[1 - \chi_h] + \delta \, (\beta \boldsymbol{B}[1] + \boldsymbol{C}[1])) \, \boldsymbol{E}_{\mathrm{f}} = (\delta \boldsymbol{C}[1] + \boldsymbol{C}[1 - \chi_h]) \, \tilde{\boldsymbol{E}}, .$$

Having deduced explicit expressions for the linear equation systems arising from (5.4), (5.11) and (5.13) we can now turn to the practical computation of χ_h. Note that $\mathfrak{P}_h \tilde{E}_h$, $E_{\mathrm{b},h}$ and $E_{\mathrm{f},h}$ are elements of $S_h^{(1)}(\Omega)$ and thus the associated coefficient vectors are in $\mathbb{R}^{(N+1)^2}$. However, $\chi_h \in S_h^{(0)}(\Omega)$ and hence $\boldsymbol{\chi} \in \mathbb{R}^{N^2}$. Thus we introduce the projection matrix that maps linear splines to constants, i.e.,

$$\boldsymbol{P}_1 := \left\{ \iint_\Omega s_i^{(0)} s_j^{(1)} \right\}_{\substack{i=1,\ldots,N^2 \\ j=1,\ldots,(N+1)^2}} \in \mathbb{R}^{N^2 \times (N+1)^2}.$$

For its stencil see Appendix C.2. Hence in practice we compute

$$\boldsymbol{\chi} = \mathrm{bool} \left(\vartheta \left| \boldsymbol{P}_1 (\boldsymbol{E}_{\mathrm{b}} - \tilde{\boldsymbol{E}}) \right| \leq \left| \boldsymbol{P}_1 (\boldsymbol{E}_{\mathrm{f}} - \tilde{\boldsymbol{E}}) \right| \right).$$

Note that the Inner Iteration was designed with the hypothesis in mind that it forms a fixed point iteration. While it has been discussed (compare Remark 1) that this is not true in general, numerous numerical tests showed, however, that in practice convergence of the Inner Iteration is quite robust. This was a key observation in the design of an appropriate stopping criterion. First, we want the Inner Iteration to terminate if the computed edge map does not change any more. However, we also want to include a safeguard that stops the iteration if the segmentation of $\tilde{E}_h$ comes to a standstill, i.e., if updating $E_{\mathrm{b},h}$ and $E_{\mathrm{f},h}$ shows no effect. Note that by construction, changes in the foreground image are reflected in the background image and vice versa. Thus including $E_{\mathrm{f},h}$ in the stopping criterion for the Inner Iteration is equivalent to using $E_{\mathrm{b},h}$ in it.

The design of the Outer Iteration was inspired by a similar fixed point assumption. However, in the course of developing Algorithm 3.1, this original motivation faded. It turned out that the quality of edge maps obtained after convergence of the Inner Iteration was usually very high. Thus the Outer Iteration rather mimics the technique of *iterative refinement* used in the context of linear equation systems (see, e.g., [DH02]). This resemblance also motivates choosing the maximal number of Outer Iterations $k_{\mathrm{out}}^{\max}$ smaller than its Inner Iteration counterpart $k_{\mathrm{in}}^{\max}$. Furthermore, these observations are also reflected in the stopping criterion for the Outer Iteration. If two successive results of the Inner Iteration do not change or if the smooth image $I_{\mathrm{s},h}$ (that is based on χ_h) stagnates we want the Outer Iteration and thus the whole Algorithm to terminate. These ideas are summarized in Algorithm 5.1.

Algorithm 5.1 Pseudo-code for Algorithm 3.1

Input: $\tilde{I}_h = \sum_{i=1}^{N^2} \tilde{I}_{h,i} s_i^{(0)}$, β, δ, ϑ
Output: χ_h, $I_{\mathrm{s},h}$

1: **Compute** an initial guess $\boldsymbol{\chi}$ using K-means clustering (Section 2.3) with $K = 2$
2: **Assemble** the matrices $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ as well as

$$\boldsymbol{M}_{\mathrm{out}} = \delta\left(\beta \boldsymbol{A}[1] + \boldsymbol{G}[1]\right) \text{ and } \boldsymbol{M}_{\mathrm{in}} = \delta\left(\beta \boldsymbol{B}[1] + \boldsymbol{C}[1]\right)$$

3: **Choose** $\mathrm{tol}_{\mathrm{out}}, \mathrm{tol}_{\mathrm{in}} \in (0,1)$ and $k_{\mathrm{out}}^{\max}, k_{\mathrm{in}}^{\min} \in \mathbb{N}_{>1}$ as well as $0 < \tau_\varepsilon \ll 1$

4: **Set** $k_{\mathrm{out}} = 0$, $r_{\mathrm{out}} = 2 \cdot \mathrm{tol}_{\mathrm{out}}$, $\tilde{\boldsymbol{I}} = \{\tilde{I}_{h,i}\}_{i=1}^{N^2}$ as well as

$$\boldsymbol{I}_{\mathrm{s}} = (0,\dots,0)^\top \in \mathbb{R}^{(N+2)^2} \text{ and } \boldsymbol{E}_{\mathrm{f}} = (0,\dots,0)^\top \in \mathbb{R}^{(N+1)^2}$$

5: **while** $r_{\mathrm{out}} > \mathrm{tol}_{\mathrm{out}}$ **and** $k_{\mathrm{out}} < k_{\mathrm{out}}^{\max}$ **do**
6:      $k_{\mathrm{out}} \leftarrow k_{\mathrm{out}} + 1$
7:      $\boldsymbol{\chi}_{\mathrm{out}} \leftarrow \boldsymbol{\chi}$ and $\boldsymbol{I}_{\mathrm{s,out}} \leftarrow \boldsymbol{I}_{\mathrm{s}}$
8:      **Solve**

$$\left(\beta \boldsymbol{A}[\chi_h] + \boldsymbol{G}[\chi_h] + \boldsymbol{M}_{\mathrm{out}}\right) \boldsymbol{I}_{\mathrm{s}} = \boldsymbol{P}_2^\top \left((\delta + \boldsymbol{\mathcal{D}}[\chi_h])\tilde{\boldsymbol{I}}\right),$$

9:      **Compute** $\tilde{E}_h = |\nabla I_{\mathrm{s},h}|$ and $\tilde{\boldsymbol{E}}$
10:     **Set** $k_{\mathrm{in}} = 0$ and $r_{\mathrm{in}} = 2 \cdot \mathrm{tol}_{\mathrm{in}}$

11:     **while** $r_{\mathrm{in}} > \mathrm{tol}_{\mathrm{in}}$ **and** $k_{\mathrm{in}} < k_{\mathrm{in}}^{\max}$ **do**
12:        $k_{\mathrm{in}} \leftarrow k_{\mathrm{in}} + 1$
13:        $\boldsymbol{\chi}_{\mathrm{in}} \leftarrow \boldsymbol{\chi}$ and $\boldsymbol{E}_{\mathrm{f,in}} \leftarrow \boldsymbol{E}_{\mathrm{f}}$
14:        **Solve**
$$\left(\beta \boldsymbol{B}[\chi_h] + \boldsymbol{C}[\chi_h] + \boldsymbol{M}_{\mathrm{in}}\right) \boldsymbol{E}_{\mathrm{b}} = \left(\delta \boldsymbol{C}[1] + \boldsymbol{C}[\chi_h]\right) \tilde{\boldsymbol{E}},$$

15:        **Solve**

$$\left(\beta \boldsymbol{B}[1 - \chi_h] + \boldsymbol{C}[1 - \chi_h] + \boldsymbol{M}_{\mathrm{in}}\right) \boldsymbol{E}_{\mathrm{f}} = \left(\delta \boldsymbol{C}[1] + \boldsymbol{C}[1 - \chi_h]\right) \tilde{\boldsymbol{E}},$$

16:        **Compute**

$$\boldsymbol{\chi} = \mathrm{bool}\left(\vartheta \left|\boldsymbol{P}_1(\boldsymbol{E}_{\mathrm{b}} - \tilde{\boldsymbol{E}})\right| \leq \left|\boldsymbol{P}_1(\boldsymbol{E}_{\mathrm{f}} - \tilde{\boldsymbol{E}})\right|\right),$$

17:        **Compute**

$$r_{\mathrm{in}} = \max |\boldsymbol{\chi}_{\mathrm{in}} - \boldsymbol{\chi}| \cdot \frac{|\boldsymbol{E}_{\mathrm{f,in}} - \boldsymbol{E}_{\mathrm{f}}|}{|\boldsymbol{E}_{\mathrm{f}}| + \tau_\varepsilon}$$

18:      **end while**
19:      **Compute**

$$r_{\mathrm{out}} = \max |\boldsymbol{\chi}_{\mathrm{out}} - \boldsymbol{\chi}| \cdot \frac{|\boldsymbol{I}_{\mathrm{s,out}} - \boldsymbol{I}_{\mathrm{s}}|}{|\boldsymbol{I}_{\mathrm{s}}| + \tau_\varepsilon}$$

20: **end while**

$$\text{Theorems 6 and 8}$$

$$u^\delta \qquad \leftarrow \qquad u_h^\delta$$

$$\text{Theorem 10} \quad \downarrow \qquad\qquad\qquad \downarrow \quad \text{Theorem 12}$$

$$u^\star \qquad \leftarrow \qquad u_h^\star$$

$$\text{Theorem 11}$$
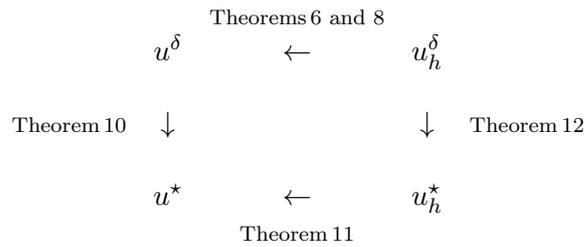
Figure 5.2: Sketch illustrating the convergence results presented in this section. The arrows "↓" indicate the limit $\delta \to 0$ while "←" means $h \to 0$.

5.4 Investigation of Vanishing Regularization Solutions

Recall that δ was introduced in the functionals F and J given by (3.1) and (3.2) respectively as a safeguard in case χ has not yet accurately converged to a characteristic function with positive measure support. Hence we always assumed that $0 < \delta \ll 1$. Thus, a natural question is whether solutions of (4.3), (4.8) and (4.63) admit a unique limit as $\delta \to 0$. The following investigation is devoted to this question and additionally addresses the asymptotic behavior of the associated finite element discretizations (5.11), (5.13) and (5.4) respectively. The convergence diagram in Figure 5.2 illustrates the procedure.

The results presented below are based on the theory given in [GR80, Chap. 1 §4, Chap. 2 §1]. For the sake of brevity we will take a more general approach in this section and introduce notation to cover both the Inner and Outer Iterations. Before we start our discussion recall Assumption 5. We presumed that $\chi_h \xrightarrow[L^\infty(\Omega)]{} \chi$ as $h \to 0$. Since χ is a characteristic function, i.e., $\mathrm{Rg}(\chi) = \{0, 1\}$, this approximation property essentially means that χ is a grid function for h small enough. Since we will also address the behavior of respective finite element discretizations in the limit of vanishing regularization we explicitly pose this condition on χ straight away.

Assumption 7. *Let* $\Omega := (0,1)^2$, $m \in \{1,2\}$ *and* $\tilde{u} \in L^2(\Omega)$. *Further, let* $\beta > 0$, $0 < \delta \ll 1$ *and assume* χ *is a characteristic function defined on a rectangular grid on* Ω *for h small enough. Suppose that* $\Omega_\chi := \mathrm{supp}(\chi) \subsetneq \Omega$ *is open, non-empty and* $\complement\bar{\Omega}_\chi$ *is open.*

Note that the premises on Ω_χ and $\complement\bar{\Omega}_\chi$ permit the definition of Sobolev spaces on these sets. Further, any open set has positive measure which allows the formulation of variational problems. In the following discussion we focus on the set Ω_χ. However, all results presented below can be applied to $\complement\bar{\Omega}_\chi$ as well by replacing χ by $1 - \chi$ in the respective forms.

Given Assumption 7 we start by introducing

$$\mathcal{G}[u, \chi] := \frac{1}{2} \int_\Omega |u - \tilde{u}|^2 \chi + \beta\chi |\nabla^m u|^2 \, d\boldsymbol{x},$$

and define

$$(5.19) \qquad \mathcal{G}_\delta[u] := \mathcal{G}[u, \chi] + \delta\mathcal{G}[u, 1].$$

Thus $\tilde{u}$ plays the role of $\tilde{E}$ in the Inner and $\tilde{I}$ in the Outer Iteration. We consider the

optimization problem

(5.20)
$$\inf_{u \in H^m(\Omega)} \mathcal{G}_\delta[u].$$

Then Theorems 1 and 3 guarantee that (5.20) has a unique solution

$$u^\delta := \arg \min_{u \in H^m(\Omega)} \mathcal{G}_\delta[u],$$

that satisfies the weak formulation of the necessary optimality conditions associated to (5.20), namely

(5.21)
$$Q_\delta[u^\delta, v] = S_\delta[v], \quad \forall v \in H^m(\Omega),$$

where $Q_\delta : H^m(\Omega) \times H^m(\Omega) \to \mathbb{R}$ is a bilinear form

$$Q_\delta[u, v] := \int_\Omega \beta(\delta + \chi)\nabla^m u : \nabla^m v + (\delta + \chi)uv \, d\boldsymbol{x},$$

with $\nabla^m u : \nabla^m v$ denoting a component-wise matrix scalar product for $m = 2$ and the standard vector scalar product for $m = 1$ and $S_\delta : H^m(\Omega) \to \mathbb{R}$ is a linear functional

$$S_\delta[v] := \int_\Omega \tilde{u}(\delta + \chi)v \, d\boldsymbol{x}.$$

We first prove a technical result that will be crucial in the following.

Lemma 7. *Given Assumption 7 consider the Hilbert spaces $H^m(\Omega_\chi)$ and $H^m(\Omega)$. Then*

1. *The form $\langle \cdot, \cdot \rangle_{H^m(\Omega_\chi)}$ defined by*

$$\langle \lambda, \mu \rangle_{H^m(\Omega_\chi)} := \int_{\Omega_\chi} \beta \nabla^m \lambda : \nabla^m \mu + \lambda \mu \, d\boldsymbol{x}, \quad \forall \lambda, \mu \in H^m(\Omega_\chi),$$

   is an inner product on $H^m(\Omega_\chi)$ that induces the norm $|\cdot|_{H^m(\Omega_\chi)} := \sqrt{\langle \cdot, \cdot \rangle_{H^m(\Omega_\chi)}}$ which is equivalent to the usual norm $\|\cdot\|_{H^m(\Omega_\chi)}$ on Ω_χ.

2. *Similarly*
$$|v|_{H^m(\Omega)}^2 := \int_\Omega \beta \left| \nabla^m v \right|^2 + v^2 d\boldsymbol{x}, \quad \forall v \in H^m(\Omega),$$

   induces a norm on $H^m(\Omega)$ which is equivalent to the usual norm $\|\cdot\|_{H^m(\Omega)}$ on Ω.

3. *There exists a bounded linear extension operator $\mathfrak{E} : H^m(\Omega_\chi) \to H^m(\Omega)$, i.e.,*

   (5.22)
$$\|\mathfrak{E}\mu\|_{H^m(\Omega)} \leq \alpha_0 |\mu|_{H^m(\Omega_\chi)}, \quad \forall \mu \in H^m(\Omega_\chi),$$

   for some $\alpha_0 > 0$.

Proof. The second claim is a consequence of Corollary 4.16 in [Ada75] which also covers the third claim. Thus there exist positive constants α_1 and α_2 such that

(5.23)
$$\alpha_1 |\mu|_{H^m(\Omega_\chi)} \leq \|\mu\|_{H^m(\Omega_\chi)} \leq \alpha_2 |\mu|_{H^m(\Omega_\chi)}, \quad \forall \mu \in H^m(\Omega_\chi).$$

Note that for $m = 1$ this inequality chain is readily established with $\alpha_1 = 1/\max\{1, \beta\}$ and $\alpha_2 = 1/\min\{1, \beta\}$. Similarly we have

$$(5.24) \qquad \alpha_3 \, |v|_{H^m(\Omega)} \leq \|v\|_{H^m(\Omega)} \leq \alpha_4 \, |v|_{H^m(\Omega)} \,, \quad \forall v \in H^m(\Omega),$$

for $\alpha_3, \alpha_4 > 0$. For the third claim observe that the corners of Ω_χ all are at right angles. Thus smooth functions in Ω_χ can be extended smoothly outside Ω_χ and consistently at all corners (i.e., horizontally and vertically yields the same results as vertically and horizontally). Hence the claim follows as in the proof of Theorem 5.4 in [Sho94, Chap. 5]. $\qquad\square$

In the following we will always endow $H^m(\Omega_\chi)$ with the introduced norm $|\cdot|_{H^m(\Omega_\chi)}$. This space will play a key role for the analysis presented below.

5.4.1 Asymptotic Behavior in the Continuum

Since we are interested in the limiting behavior of u^δ as $\delta \to 0$ we start by looking at $Q_{\delta=0}$. Note that while Q_δ is defined on $H^m(\Omega)$ the natural domain of $Q_{\delta=0}$ are Sobolev functions on Ω_χ. Thus, heuristically speaking, the reference spaces change in the limiting process. This makes a straight forward analysis of (5.21) very difficult. Hence we employ considerations from *duality theory* (see, e.g., [ET99]). We decouple terms in (5.21) that depend on χ from those that do not in order to obtain a *primal-dual formulation* of (5.21). Thus using the previously introduced spaces we define bilinear forms $q : H^m(\Omega) \times H^m(\Omega) \to \mathbb{R}$ and $r : H^m(\Omega) \times H^m(\Omega_\chi) \to \mathbb{R}$ given by

$$q[u, v] := \int_\Omega \beta \nabla^m u : \nabla^m v + uv \, d\boldsymbol{x},$$

and

$$r[u, \lambda] = \int_\Omega (\beta \nabla^m u : \nabla^m \lambda + u\lambda) \, \chi \, d\boldsymbol{x},$$

respectively. Further, let $z : H^m(\Omega_\chi) \times H^m(\Omega_\chi) \to \mathbb{R}$ be defined by $z[\lambda, \mu] := \langle \lambda, \mu \rangle_{H^m(\Omega_\chi)}$. Note that q and r are continuous since for any $u, v \in H^m(\Omega)$ and $\lambda \in H^m(\Omega_\chi)$ we have

$$(5.25) \qquad |q[u, v]| \leq (\beta + 1) \, \|u\|_{H^m(\Omega)} \, \|v\|_{H^m(\Omega)} \,,$$

and

$$(5.26) \qquad |r[u, \lambda]| \leq (\beta + 1) \, \|u\|_{H^m(\Omega)} \, |\lambda|_{H^m(\Omega_\chi)} \,.$$

Obviously z is continuous as well (by Lemma 7). Further, we introduce the functionals $s_1 : H^m(\Omega) \to \mathbb{R}$ and $s_\chi : H^m(\Omega_\chi) \to \mathbb{R}$ defined by

$$s_1[v] := \int_\Omega \tilde{u} v \, d\boldsymbol{x}, \quad \text{and} \quad s_\chi[\mu] := \int_{\Omega_\chi} \tilde{u} \mu \, d\boldsymbol{x},$$

which are also continuous

$$(5.27) \qquad |s_1[v]| \leq \|\tilde{u}\|_{L^2(\Omega)} \, \|v\|_{H^m(\Omega)} \,, \quad \text{and} \quad |s_\chi[\mu]| \leq \|\tilde{u}\|_{L^2(\Omega)} \, |\mu|_{H^m(\Omega_\chi)} \,.$$

Then a primal dual formulation of (5.21) is established as follows.

Theorem 9. *Given Assumption 7 there exists a unique $(u^\delta, \lambda^\delta) \in H^m(\Omega) \times H^m(\Omega_\chi)$ such that*

$$q[u^\delta, v] + r[v, \lambda^\delta] = s_1[v], \qquad \forall v \in H^m(\Omega), \tag{5.28a}$$

$$r[u^\delta, \mu] - \delta z[\lambda^\delta, \mu] = s_\chi[\mu], \qquad \forall \mu \in H^m(\Omega_\chi), \tag{5.28b}$$

where $u^\delta \in H^m(\Omega)$ denote the unique solution of (5.21).

Proof. Let $\mu \in H^m(\Omega_\chi)$. Then by Lemma 7, every $\mu \in H^m(\Omega_\chi)$ permits an extension $\mathfrak{E}\mu = \mu_e \in H^m(\Omega)$. Clearly,

$$\begin{aligned} r[v, \mu_e] &= r[v, \mu], \quad \forall v \in H^m(\Omega), \\ z[\lambda, \mu_e] &= z[\lambda, \mu], \quad \forall \lambda \in H^m(\Omega_\chi), \\ s_\chi[\mu_e] &= s_\chi[\mu]. \end{aligned} \tag{5.29}$$

Thus (5.28b) can be extended in $\mu_e \in H^m(\Omega)$. Hence let $v = \mu_e$ in (5.28a). Then by multiplying (5.28a) by δ and adding both equations of (5.28), λ^δ is eliminated and (5.21) is obtained. Let u^δ be given by (5.21). Then (5.28b) can be rewritten such that

$$\delta z[\lambda^\delta, \mu] = \tilde{r}[\mu], \quad \forall \mu \in H^m(\Omega_\chi), \tag{5.30}$$

where $\tilde{r}[\mu] := r[u^\delta, \mu] - s_\chi[\mu]$. According to (5.26) and (5.27), r and s_χ are bounded; thus, $\tilde{r}$ is continuous. By Lemma 7, z is bounded and coercive on $H^m(\Omega_\chi)$. Hence according to the Lax–Milgram Lemma (Appendix B.2), let λ^δ be given by (5.30). Suppose $(\hat{u}^\delta, \hat{\lambda}^\delta)$ is another solution to (5.28). Then by above any $\hat{u}^\delta$ must solve (5.21), so $\hat{u}^\delta = u^\delta$. Thus, $r[\hat{u}^\delta, \mu] = r[u^\delta, \mu]$ for all $\mu \in H^m(\Omega_\chi)$ which in (5.30) yields $\hat{\lambda}^\delta = \lambda^\delta$. Hence uniqueness of $(u^\delta, \lambda^\delta)$ follows. $\qquad\square$

Having established the primal-dual formulation (5.28) we now need to characterize a possible limit $u^\star$ of u^δ as $\delta \to 0$. A natural candidate for $u^\star$ is the solution (if existent) of (5.28) with $\delta = 0$ such that δz vanishes. A less heuristic derivation follows considerations from optimization theory. Thus suppose the limit $u^\star$ exists. With a closer look at (5.19) we might expect that $u^\star$ solves

$$\min \mathcal{G}[u, 1] \quad \text{subject to} \quad \frac{\delta \mathcal{G}}{\delta u}[u, \chi; v] = 0, \quad \forall v \in C^\infty(\bar{\Omega}), \tag{5.31}$$

where $\frac{\delta \mathcal{G}}{\delta u}[u, \chi; v]$ denotes the Gâteaux derivative (see Appendix B.2) of $\mathcal{G}$ in an arbitrary direction $v \in C^\infty(\bar{\Omega})$. The standard approach for solving such a constrained optimization problem is to consider the associated Lagrange functional (compare, e.g., [GR80, Chap. 1, Sec. 4.2]). Thus we start by computing

$$\begin{aligned} \frac{\delta \mathcal{G}}{\delta u}[u, \chi; v] &= \frac{d}{ds} \left(\frac{1}{2} \int_\Omega \left(|u + sv - \tilde{u}|^2 + \beta |\nabla^m(u + sv)|^2 \right) \chi \, d\boldsymbol{x} \right) \\ &= \int_\Omega \left((u - \tilde{u})v + \beta \nabla^m u : \nabla^m v \right) \chi \, d\boldsymbol{x}, \quad \forall v \in C^\infty(\bar{\Omega}). \end{aligned}$$

By Lemma 7, every $\lambda \in H^m(\Omega_\chi)$ can be approximated arbitrarily well in $H^m(\Omega_\chi)$ by some $v \in C^\infty(\bar{\Omega})$. Hence

$$\frac{\delta \mathcal{G}}{\delta u}[u, \chi; \lambda] = \int_\Omega ((u - \tilde{u})v\lambda + \beta \nabla^m u : \nabla^m \lambda) \chi \, d\boldsymbol{x}, \quad \forall \lambda \in H^m(\Omega_\chi).$$

Then the Lagrangian associated to (5.31) is given by

$$\mathfrak{L}[u, \lambda] := \mathcal{G}[u, 1] + \int_\Omega ((u - \tilde{u})\lambda + \beta \nabla^m u : \nabla^m \lambda) \chi \, d\boldsymbol{x}, \quad u \in H^m(\Omega), \lambda \in H^m(\Omega_\chi),$$

and thus we (formally) obtain the optimality conditions

$$\frac{\delta \mathfrak{L}}{\delta u}[u, \lambda; v] = \int_\Omega (u - \tilde{u})v + \beta \nabla^m u : \nabla^m v + \int_\Omega (v\lambda + \beta \nabla^m v : \nabla^m \lambda) \chi \, d\boldsymbol{x}, \quad \forall v \in H^m(\Omega),$$

and

$$\frac{\delta \mathfrak{L}}{\delta \lambda}[u, \lambda; \mu] = \int_\Omega ((u - \tilde{u})\mu + \beta \nabla^m u : \nabla^m \mu) \chi \, d\boldsymbol{x}, \quad \forall \mu \in H^m(\Omega_\chi).$$

Then using the notation introduced above the optimality system

$$\begin{cases} \dfrac{\delta \mathfrak{L}}{\delta u}[u, \lambda; v] = 0, & \forall v \in H^m(\Omega), \\ \dfrac{\delta \mathfrak{L}}{\delta \lambda}[u, \lambda; \mu] = 0, & \forall \mu \in H^m(\Omega_\chi), \end{cases}$$

leads indeed to the following saddle point problem. Find $(u^\star, \lambda^\star) \in H^m(\Omega) \times H^m(\Omega_\chi)$ satisfying

(5.32a) $\qquad q[u^\star, v] + r[v, \lambda^\star] = s_1[v], \qquad\qquad \forall v \in H^m(\Omega),$

(5.32b) $\qquad r[u^\star, \mu] \qquad\quad = s_\chi[\mu], \qquad\qquad \forall \mu \in H^m(\Omega_\chi).$

The Theorem below proves the conjecture that $(u^\star, \lambda^\star)$ is the actual limit of $(u^\delta, \lambda^\delta)$ as $\delta \to 0$.

Theorem 10. *Given Assumption 7 let $(u^\delta, \lambda^\delta) \in H^m(\Omega) \times H^m(\Omega_\chi)$ be the unique solution to (5.28). Then (5.32) has a unique solution $(u^\star, \lambda^\star) \in H^m(\Omega) \times H^m(\Omega_\chi)$ and $u^\delta \xrightarrow[H^m(\Omega)]{} u^\star$ as well as $\lambda^\delta \xrightarrow[H^m(\Omega_\chi)]{} \lambda^\star$ as $\delta \to 0$.*

Proof. Theorem 9 guarantees existence and uniqueness of $(u^\delta, \lambda^\delta) \in H^m(\Omega) \times H^m(\Omega_\chi)$. We show that the Conditions [GR80, (4.9), (4.25), (4.31)] of Theorem 4.3 in [GR80, Chap. 1, Sec. 4.3] are met. Thus we start by proving that

(5.33) $$\inf_{\mu \in H^m(\Omega_\chi)} \sup_{v \in H^m(\Omega)} \frac{r[v, \mu]}{\|v\|_{H^m(\Omega)} |\mu|_{H^m(\Omega_\chi)}} \geq \alpha_r,$$

for some positive real number α_r. By Lemma 7 any $\mu \in H^m(\Omega_\chi)$ permits an extension

$\mathfrak{E}\mu = \mu_{\mathrm{e}} \in H^m(\Omega)$. Then by (5.22) and (5.29) we get

(5.34)
$$\sup_{v \in H^m(\Omega)} \frac{r[v, \mu]}{\|v\|_{H^m(\Omega)}} = \frac{r[v, \mu_{\mathrm{e}}]}{\|v\|_{H^m(\Omega)}} \geq \frac{r[\mu_{\mathrm{e}}, \mu_{\mathrm{e}}]}{\|\mu_{\mathrm{e}}\|_{H^m(\Omega)}} = \frac{r[\mu, \mu]}{\|\mu_{\mathrm{e}}\|_{H^m(\Omega)}}$$
$$\geq \frac{|\mu|^2_{H^m(\Omega_\chi)}}{\alpha_0 \, |\mu|_{H^m(\Omega_\chi)}} = \frac{1}{\alpha_0} \, |\mu|_{H^m(\Omega_\chi)} \, .$$

Hence (5.33) holds with $\alpha_r = 1/\alpha_0$ and thus Condition [GR80, (4.9)] is established. Note further that since $z[\mu, \mu] = |\mu|^2_{H^m(\Omega_\chi)}$ the operator $\mathcal{Z} \in \mathcal{L}(H^m(\Omega_\chi), H^m(\Omega_\chi)')$ associated to z by

$$\langle \mathcal{Z}\lambda, \mu \rangle_{H^m(\Omega_\chi)} = z[\lambda, \mu], \quad \forall \lambda, \mu \in H^m(\Omega_\chi),$$

is the identity and thus Condition [GR80, (4.25)] follows. Let similarly $\mathfrak{R} \in \mathcal{L}(H^m(\Omega), H^m(\Omega_\chi)')$ be the operator associated to r. Thus $\mathfrak{R}$ is the continuous embedding of $H^m(\Omega)$ into $H^m(\Omega_\chi)'$ and hence

(5.35)
$$\langle \mathcal{Z}^{-1}\mathfrak{R}v, \mathfrak{R}v \rangle_{H^m(\Omega)} = q[v, v], \quad \forall v \in H^m(\Omega).$$

Note that $q[v, v] = |v|^2_{H^m(\Omega)}$ and by Lemma 7 $|\cdot|_{H^m(\Omega)}$ is equivalent to $\|\cdot\|_{H^m(\Omega)}$ on $H^m(\Omega)$. Thus by (5.24) we have

(5.36)
$$q[v, v] \geq \frac{1}{\alpha_4} \|v\|^2_{H^m(\Omega)}, \quad \forall v \in H^m(\Omega),$$

and hence the final Condition [GR80, (4.31)] of Theorem 4.3 in [GR80, Chap. 1, Sec. 4.3] is met which proves the claim. $\qquad\square$

Note that $u_h^\delta \to u^\delta$ has been shown in Theorems 6 and 8. Thus in order to complete the convergence sketch of Figure 5.2 it remains to prove that $u_h^\delta \to u_h^\star$ and $u_h^\star \to u^\star$.

5.4.2 Asymptotic Behavior of Finite Element Discretizations

Recall that by Assumption 7, χ is a grid function for h small enough. Thus for some h we have $\chi_h = \chi$. Hence to simplify the argumentation below suppose h is sufficiently small so that we have $\chi_h = \chi$ straight away. The second item of Assumption 5 governs data approximation. Recall that we presumed that $\tilde{I}_h \xrightarrow[L^\infty(\Omega)]{} \tilde{I}$ which does not pose any restriction on the raw data $\tilde{I}$. Under this condition, however, proving convergence of finite element discretizations of (5.28) and (5.32) requires profound adaptations of the classical theory given in [GR80] and makes the argumentation substantially more complex. The recent work [KKP12] develops the necessary technical means and presents a rigorous analysis of finite element discretizations that respect the pixelwise averaged data representation characteristic to digital images. However, given that the focus of the present work is not on numerical analysis of finite elements, we decided to follow the standard approach of [GR80] by assuming that $\tilde{I}_h = \tilde{I}$. In other words, we suppose that the raw image is piecewise constant. We summarize these considerations in the following Assumption.

Assumption 8. *Suppose h is sufficiently small so that $\chi_h = \chi$ and assume further that $\tilde{u}_h = \tilde{u}$.*

Now we can start by formulating a finite element discretization of (5.32) as follows. Find $(u_h^\star, \lambda_h^\star) \in S_h^{(m)}(\Omega) \times S_h^{(m)}(\Omega_\chi)$ such that

$$(5.37a) \qquad q[u_h^\star, v_h] + r[v_h, \lambda_h^\star] = s_1[v_h], \forall v_h S_h^{(m)}(\Omega),$$

$$(5.37b) \qquad r[u_h^\star, \mu_h] = s_\chi[\mu_h], \forall \mu_h \in S_h^{(m)}(\Omega_\chi).$$

The following Theorem shows that (5.37) indeed approximates (5.32).

Theorem 11. *Given Assumptions 7 and 8 let $(u^\star, \lambda^\star) \in H^m(\Omega) \times H^m(\Omega_\chi)$ denote the unique solution to (5.32). Then (5.37) has a unique solution $(u_h^\star, \lambda_h^\star) \in S_h^{(m)}(\Omega) \times S_h^{(m)}(\Omega_\chi)$ and $u_h^\star \xrightarrow[H^m(\Omega)]{} u^\star$ as well as $\lambda_h^\star \xrightarrow[H^m(\Omega_\chi)]{} \lambda^\star$ as $h \to 0$.*

Proof. Theorem 10 guarantees existence and uniqueness of $(u^\star, \lambda^\star) \in H^m(\Omega) \times H^m(\Omega_\chi)$. We show that the Conditions [GR80, (1.10), (1.12)] of Theorem 1.1.2°) in [GR80, Chap. 2, Sec. 1.1] are met. Note that $S_h^{(m)}(\Omega) \subset H^m(\Omega)$ and $S_h^{(m)}(\Omega_\chi) \subset H^m(\Omega_\chi)$. Thus (5.34) immediately implies

$$\sup_{v_h \in S_h^{(m)}(\Omega)} \frac{r[v_h, \mu_h]}{\|v_h\|_{H^m(\Omega)}} \geq \frac{1}{\alpha_0} |\mu_h|_{H^m(\Omega_\chi)}, \quad \forall \mu_h \in S_h^{(m)}(\Omega_\chi),$$

hence [GR80, (1.12)] holds. Similarly (5.36) yields

$$(5.38) \qquad q[v_h, v_h] \geq \frac{1}{\alpha_4} \|v_h\|_{H^m(\Omega)}^2, \quad \forall v_h \in S_h^{(m)}(\Omega),$$

and thus [GR80, (1.10)] is satisfied. Then Theorem 1.1.2°) in [GR80, Chap. 2, Sec. 1.1] proves the claim. $\square$

Similarly, we set up a finite element discretization of (5.28) as follows. Find $(u_h^\delta, \lambda_h^\delta) \in S_h^{(m)}(\Omega) \times S_h^{(m)}(\Omega_\chi)$ such that

$$(5.39a) \qquad q[u_h^\delta, v_h] + r[v_h, \lambda_h^\delta] = s_1[v_h], \forall v_h \in S_h^{(m)}(\Omega),$$

$$(5.39b) \qquad r[u_h^\delta, \mu_h] - \delta z[\lambda_h^\delta, \mu_h] = s_\chi[\mu_h], \forall \mu_h \in S_h^{(m)}(\Omega_\chi).$$

Then the result below proves that for small δ (5.39) and (5.37) approximate (5.28) and (5.32) respectively which finally completes the convergence sketch of Figure 5.2.

Theorem 12. *Given Assumptions 7 and 8 let $(u_h^\star, \lambda_h^\star) \in S_h^{(m)}(\Omega) \times S_h^{(m)}(\Omega_\chi)$ denote the unique solution to (5.37). Then (5.39) also as a unique solution $(u_h^\delta, \lambda_h^\delta) \in S_h^{(m)}(\Omega) \times S_h^{(m)}(\Omega_\chi)$ that satisfies $u_h^\delta \xrightarrow[H^m(\Omega)]{} u_h^\star$ and $\lambda_h^\delta \xrightarrow[H^m(\Omega_\chi)]{} \lambda_h^\star$ as $\delta \to 0$.*

Proof. Theorem 11 guarantees existence and uniqueness of $(u_h^\star, \lambda_h^\star) \in S_h^{(m)}(\Omega) \times S_h^{(m)}(\Omega_\chi)$. The rest of the claim is an immediate consequence of Theorem 10 which shows that the Conditions [GR80, (1.12), (1.24), (1.27)] of Theorem 1.3 in [GR80, Chap. 2, Sec. 1.2] are met. Since $S_h^{(m)}(\Omega) \subset H^m(\Omega)$ and $S_h^{(m)}(\Omega_\chi) \subset H^m(\Omega_\chi)$ relation (5.35) holds for any $v_h \in S_h^{(m)}(\Omega)$. Hence [GR80, (1.24), (1.27)] are satisfied. Further, (5.38) implies that [GR80, (1.12)] must hold as well. Hence all conditions of Theorem 1.3 in [GR80, Chap. 2, Sec. 1.2] are met which proves the claim. □

The next chapter presents results of Algorithm 5.1 and discusses its performance compared to existing approaches.

Numerical Results

The results shown in this chapter have been computed on a Dell Optiplex 745 with 8GB of RAM. The operating system used was openSUSE Linux 12.1 based on a 64bit architecture using kernel 3.5.2-39. All codes have been written in MATLAB [MAT10] and Python [vRD12] making extensive use of the packages NumPy and SciPy [JOP$^+$]. In a slight abuse of notation we will refer to all quantities by their continuum notation, i.e., we will use χ instead of χ_h or $\boldsymbol{\chi}$ to denote a binary edge map. However, all images depicted below are of course digital grayscale images given as square $N \times N$ arrays of pixels. In the following we discuss the performance of the segmentation approach introduced in practice. We start by considering an artificial image and compare results of Algorithm 5.1 to existing methods. Later in the chapter we will focus on magnetic resonance (MR) images taken from a dynamic contrast enhanced (DCE) sequence.

6.1 Choosing Parameters

We return to the artificial image depicted in Figure 2.2 that was already used throughout Chapter 2. The resolution of the image is $N = 256$ and its intensities are normalized to be in the range $[0, 1]$. Figure 3.5 shows the result of Algorithm 5.1 for $\beta = 5$, $\delta = 0.2$ and $\vartheta = 1$. Note that β is a smoothing parameter, so large values of β penalize $\left|\nabla^2 I_{\mathrm{s}}\right|^2$ on $\mathrm{supp}(\chi)$ and thus induce I_{s} to be smooth on connected components of the support of the binary edge map. Since the image considered here is artificial and free of noise we choose β small; in the presence of noise, however, β has to be increased significantly (which is addressed in the next section). Thus an adequate choice of β is crucial for the performance of Algorithm 5.1. Note, however, that the approach presented is not particularly sensitive to β. For instance, setting $\beta = 15$ and $\delta = 0.2$ for the artificial image of Figure 2.2 yields a result that is visually identical to the one shown in Figure 3.5. Thus only greatly ill-chosen values for β deteriorate the computed edge map.

Recall that δ was introduced as a safeguard for the case that initial guesses for the edge map χ are inaccurate. Hence δ should always be significantly smaller than β. It turned out that $\delta \in [0.001, 1]$ yields good results for most images. Note, however, that Algorithm 5.1 is rather stable with respect to δ, e.g., with $\delta = 0.5$ and $\beta = 5$ for the artificial image of Figure 2.2 the computed results are visually identical. We only observed negative effects if $\delta \approx \beta$. However, given the role of δ in modeling the functionals F and J given by (3.1) and (3.2), respectively, this behavior was to be expected.

As pointed out above ϑ may be seen as an edge thickness parameter. Low values of ϑ give thin edges whereas larger values induce the zero regions of χ to grow, yielding thicker edges. Hence, especially compared to β or δ, rather small changes in ϑ have a discernible impact on the final result obtained by Algorithm 5.1. An important observation in this context is the dependence of ϑ on the resolution of the considered digital image. Of course,

k_{out}	k_{in}	r_{in}	r_{out}
1	1	1	–
1	2	0	1
2	1	0.0098585	0.0025984
3	1	0	0

k_{out}	k_{in}	r_{in}	r_{out}
1	1	1	–
1	2	0	1
2	1	0.0098585	–
2	2	0	0.0025984
3	1	0	0

Table 6.1: The convergence history of Algorithm 5.1 given the artificial image depicted in Figure 2.2. Shown is the iteration progression for $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 2$ (left) and $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 12$ (right).

all parameters depend implicitly on the size of the image $\tilde{I}$, e.g., $\beta = 5$ may be too large for the same artificial image at the lower resolution $N = 32$. However, due to the prominent role of ϑ in the computation of χ the influence of the image resolution on ϑ is immediately visible. Especially low values of ϑ require sufficiently large values of N (or equivalently h small enough). Note that this can be remedied to some extent by always using $\vartheta = 1$ in the first run of the Inner Iteration, then reducing ϑ to the desired value for subsequent cycles. However, in general setting $\vartheta = 1$ turned out to be a valid choice for most images and thus this value was used as a default value.

It was mentioned above that the Outer Iteration rather resembles an iterative refinement strategy than a fixed point iteration. Thus we always choose $k_{\text{out}}^{\max} \leq k_{\text{in}}^{\max}$. Note that for most images $k_{\text{out}}^{\max} = 2$ already provides visually good results even if the error r_{out} is not (yet) below tol_{out}. We have seen that continuing to run Algorithm 5.1 to numerical convergence of the Outer Iteration, i.e . $r_{\text{out}} < \text{tol}_{\text{out}}$, does not significantly improve the quality of the final result in many cases. Thus for all simulations presented below we chose $k_{\text{out}}^{\max}$ between 2 and 5. The maximal number of steps of the Inner Iteration is chosen as follows. Note that every step of the Inner Iteration involves solving two linear equation systems. Thus large values of $k_{\text{in}}^{\max}$ may result in high computational cost if convergence is slow. On the other hand picking $k_{\text{in}}^{\max}$ very small may increase performance at the expense of quality if the computed edge map is not sufficiently close to its limit. Thus there is a trade off between performance and quality. In practice, values between 10 and 20 for $k_{\text{in}}^{\max}$ provided good edge maps without being computationally very expensive. Note that we have observed that the first run of the Inner Iteration may not converge if the initial guess for χ was poor. However, subsequent runs of the Inner Iteration tend to converge within the first 5-6 steps. If $r_{\text{in}} \geq \text{tol}_{\text{in}}$ after 10 steps then the convergence is usually slow and generally the converged edge map does not significantly differ from χ after 10 steps. Note that, of course, the choice of $k_{\text{in}}^{\max}$ and $k_{\text{out}}^{\max}$ is tightly connected to the specific values of the stopping tolerances tol_{in} and tol_{out} respectively. While the value of tol_{out} is less critical since $k_{\text{out}}^{\max}$ is usually selected very low anyway, tol_{in} has to be chosen more carefully. It turned out that $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 2$ provided satisfactory performance in practice. As a rule of thumb we observed that once r_{in} is lower than $1.0e - 2$ usually within the next two or three steps a fixed point is reached, i.e., χ becomes stationary such that $r_{\text{in}} = 0$.

To further illustrate these considerations Table 6.1 shows the convergence history for the result shown in Figure 3.5 using first $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 2$ (left table) then $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 12$ (right table). Note that indeed one further Inner Iteration in the second run of the Outer Iteration suffices to get $r_{\text{in}} = 0$. However, the result obtained using stricter error tolerances is visually identical. Note further that solving (5.4)
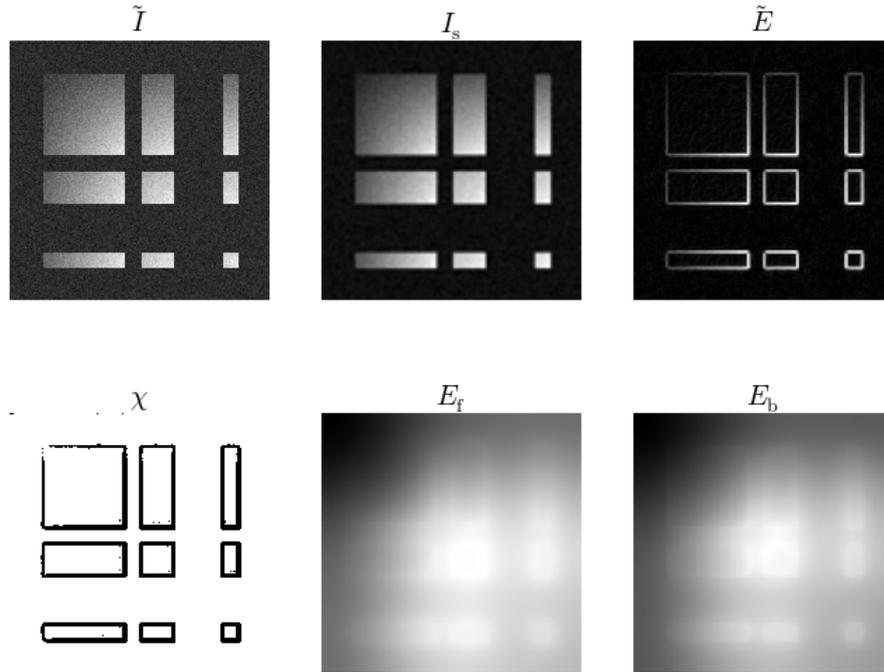
Figure 6.1: Result of Algorithm 5.1 for an artificial image corrupted by 5% additive Gaussian white noise using the parameter values $\beta = 5e4$, $\delta = 0.4$ and $\vartheta = 1$.

takes about three seconds in MATLAB (on the above specified system) while the solution of (5.11) and (5.13) is computed within around one second each. This difference is mainly due to the more complex density pattern of the coefficient matrix in (5.4) (compare the stencils given in Appendix C.2). The elaborate nonzero structure of the bending matrix $\boldsymbol{A}[\chi_h]$ makes the solution of (5.4) computationally more expensive than solving (5.11) or (5.13). Moreover, the size of the system set up to compute I_s is larger.

Note that in the course of the development of Algorithm 5.1 various parameter setups have been thoroughly tested. For instance, it turned out that determining β and δ independently for the Inner and Outer Iteration has little to no effect on the quality of results or convergence speed. Moreover, an adaptive strategy for determining $k_{\text{in}}^{\text{max}}$ was tested. The idea was to choose rather large values for both $k_{\text{out}}^{\text{max}}$ and $k_{\text{in}}^{\text{max}}$, e.g., $k_{\text{out}}^{\text{max}} = 20$ and $k_{\text{in}}^{\text{max}} = 50$, and successively reduce $k_{\text{in}}^{\text{max}}$ in subsequent steps of the Outer Iteration. Since this strategy did not noticeably improve results but led to a substantial drop in performance for some images we decided against it. Note that relaxation strategies for E_{b}, E_{f} and I_s have been investigated as well. None of which, however, improved results or performance of Algorithm 5.1 significantly.

k_{out}	k_{in}	r_{in}	r_{out}
1	1	1	–
1	2	0.062202	–
1	3	0.0001253	1
2	1	3.8203e-06	3.0418e-07

k_{out}	k_{in}	r_{in}	r_{out}
1	1	1	–
1	2	0.062202	–
1	3	0.0001253	–
1	4	0	1
2	1	9.3604e-07	–
2	2	0.0019681	–
2	3	0	3.0418e-07
3	1	0	0

Table 6.2: The convergence history of Algorithm 5.1 for the artificial image corrupted by 5% additive Gaussian white noise depicted in Figure 2.8. Shown is the iteration progression for $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 2$ (left) and $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 12$ (right).

6.2 The Noisy Artificial Image

In order to put better the performance of our approach into the context of existing methods we focus now on a more challenging example. We consider the artificial image introduced in Chapter 2 corrupted by 5% additive Gaussian white noise (as for instance depicted in Figure 2.8). Figure 6.1 shows the result of Algorithm 5.1 for this image. Due to the presence of noise we had to increase significantly the value of β (compared to the noise free case) so that I_{s} and hence $\tilde{E} = |\nabla I_{\text{s}}|$ become sufficiently smooth on the support of χ. Note that depending on the quality of the initial guess for χ, a too small value for the safeguard δ may deteriorate results for such large values of β. Algorithm 5.1 again proves its stability with respect to β and δ, i.e., mild variations in both parameters do not change the result visibly. Only a significant drop in β, e.g., $\beta \leq 1e3$, reduces the smoothness of I_{s} and thus leads to visible noise artifacts in χ. As addressed above using low stopping tolerances proved to be beneficial in this case. Table 6.2 shows the convergence history of Algorithm 5.1 for $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 2$ and $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 12$. Note that stricter stopping tolerances led to significantly more iterations while having no visible effect on the result.

We start our comparison of Figure 6.1 to results obtained by existing methods by first considering the TV denoised reconstruction seen in Panel (c) of Figure 2.9 (for $\nu = 0.4$). The corresponding Chambolle edges $|\boldsymbol{p}|_1$ and $|\boldsymbol{p}|_\infty$ are depicted in Panels (a) and (c) of Figure 2.10 respectively. As pointed out in Section 2.4 in contrast to the edge map computed by Algorithm 5.1 neither $|\boldsymbol{p}|_1$ nor $|\boldsymbol{p}|_\infty$ are binary. Thus thresholding would be necessary in order for an evaluation of χ with respect to $|\boldsymbol{p}|_1$ or $|\boldsymbol{p}|_\infty$ to make sense. However, to avoid introducing an additional layer of processing and thus potential corruption of results we rely on the non-binary $\tilde{E}$ for comparison instead. Note that in contrast to the Chambolle edges $\tilde{E}$ manifests less background noise which makes $\tilde{E}$ a potentially better candidate for the extraction of a binary edge map. However, especially in the upper left corner $\tilde{E}$ exhibits some very weak edges that are better visible in $|\boldsymbol{p}|_1$ or $|\boldsymbol{p}|_\infty$. Note further that I_{TV} may look "cleaner" than I_{s}, however, the TV denoised image shows staircasing artifacts that are not present in I_{s}. The smooth image computed by Algorithm 5.1 thus better reflects the gradual intensity variations seen in the clean original image.

Next, we focus on an Ambrosio–Tortorelli segmentation of the same noisy image. Figure 6.2 presents the results. In order to account for the noise corruption in the raw image

(a)   (b)
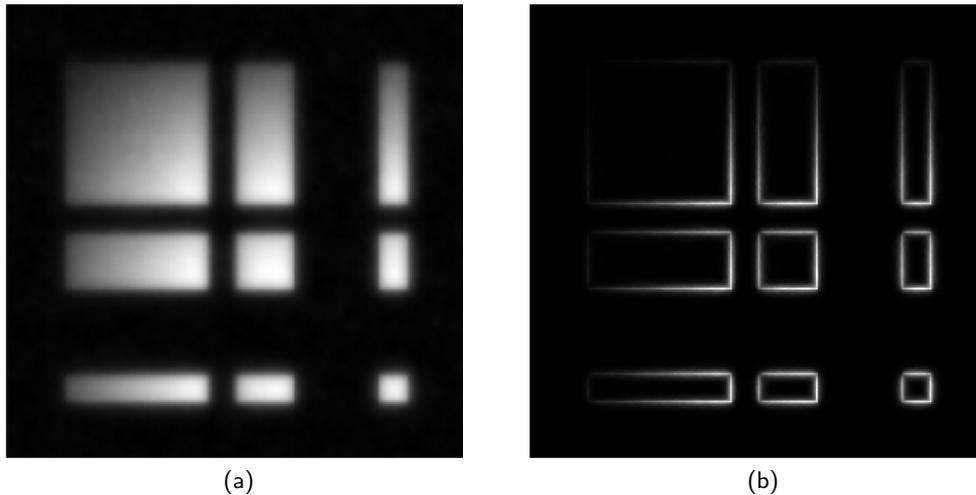
Figure 6.2: Ambrosio–Tortorelli segmentation of the artificial image corrupted by 5% additive Gaussian white noise depicted in Figure 2.8. Shown is the computed reconstruction (a) and phase function (b) for the parameters $\varepsilon = 0.1$ and $\kappa = 10$.

we set $\kappa = 10$ in J_{AT} given by (2.10). This, however, introduces significant blur in the computed reconstruction I_{AT}. Nevertheless, compared to the TV denoising result, I_{AT} better resembles the smooth shades of the original image. However, smoothness of intensity variations in I_{AT} is achieved at the expense of sharp edges. In comparison I_{s} exhibits much less blurring while still preserving edges. For reasons identical to those stated in the case of Chambolle edges we refrain from postprocessing the computed Ambrosio–Tortorelli phase function ψ_{ε} to make it comparable to χ. We rather examine it in contrast to the fuzzy edge map $\tilde{E}$. Thus for better legibility Figure 6.2 does not show ψ_{ε} but $1 - \psi_{\varepsilon}$ such that edges appear as bright lines. As opposed to the Ambrosio–Tortorelli reconstruction ψ_{ε} does not exhibit any obvious blurring. However, since I_{AT} suffers from smoothing the phase function has very weak edges as darker nonzero regions of I_{AT} fade into the background. In contrast $\tilde{E}$ shows much more pronounced edges in these areas (especially in the upper left corner of the image).

For assessing the quality of χ we first compare it also to a binary edge map computed by Canny's edge detector. Figure 6.3 shows Canny edge maps of the same noise corrupted artificial image with and without preprocessing the raw image (Panels (b) and (a) respectively). Note that both edge maps were obtained by MATLAB's Canny edge detector using automatic thresholding. The result computed without any preprocessing shown in Panel (a) clearly suffers from pronounced noise artifacts. Applying Canny's edge detector to the TV denoised image seen in Panel (c) of Figure 2.9 (for $\nu = 0.4$) improves the quality of the edge map but manifests staircasing artifacts (Panel (b)). In contrast, edges seen in χ computed by Algorithm 5.1 are straight lines that show almost no noise corruption at all. Furthermore, the depicted result was obtained without using any preprocessing strategy for the raw image. Additionally, contrary to Canny's edge detector, Algorithm 5.1 is quite robust with respect to the choice of β and δ.

To further elucidate the behavior of our approach we return again to the Canny edge maps of the clean artificial image shown in Figure 2.4. Note carefully that even in the noise free case small variations in the threshold τ cause significant changes in the computed edge
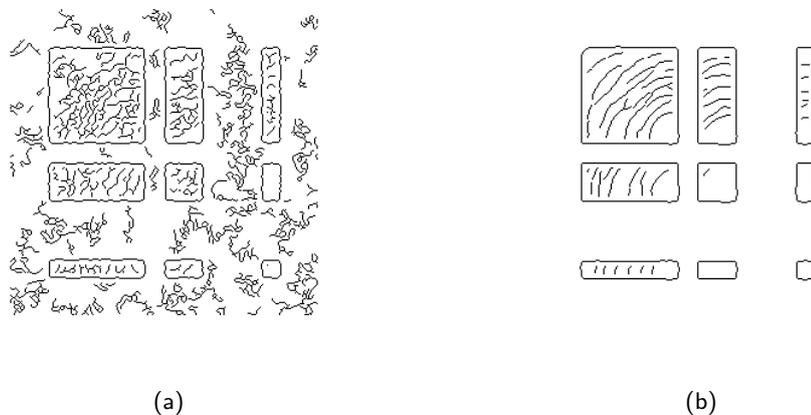
Figure 6.3: Canny edge maps of the artificial image corrupted by 5% additive Gaussian white noise depicted in Figure 2.8. Panel (a) shows a Canny edge map of the raw image without any preprocessing, Panel (b) depicts the result of Canny's detector applied to the TV denoised image seen in Panel (c) of Figure 2.9.
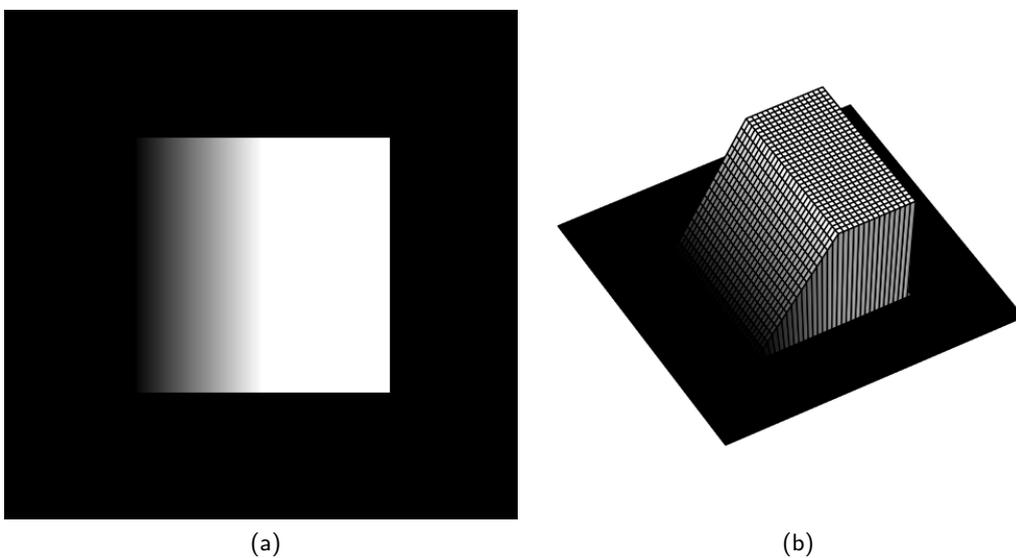
Figure 6.4: An artificial image that smoothly fades into the background represented as a gray scale map (a) and as a surface (b).
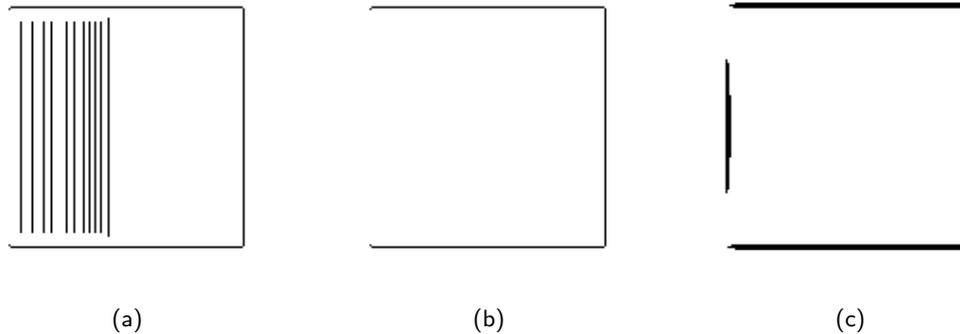
Figure 6.5: Edge maps of the artificial image depicted in Panel (a) of Figure 6.4. Shown are Canny edge maps obtained by using MATLAB's automatic thresholding (a) and by choosing the threshold value manually (b). Note that the difference in threshold values that yield the edge maps of Panel (a) and (b) is of the order of less than one tenth of a percent. Panel (c) shows an edge map computed by Algorithm 5.1 using $\beta = 2.5e3$, $\delta = 0.1$ and $\vartheta = 1$.

maps. Apparently, smooth intensity variations in the raw image prove to be challenging for Canny's edge detector. We analyze this phenomenon in more detail by considering the image depicted in Panel (a) of Figure 6.4. Note that the left side of this shaded square smoothly fades into the background which is further illustrated in the surface plot seen in Panel (b) of the same Figure. Hence the left edge of this image does not correspond to a jump but rather a sharp bend in the surface plot.

This particular kind of edge may pose problems for formulations that are solely based on locally maximizing the gradient magnitude, such as Canny's edge detector. Panels (a) and (b) of Figure 6.5 show Canny edge maps of the image depicted in Panel (a) of Figure 6.4 for different threshold values. In comparison Panel (c) of the same Figure shows an edge map computed by Algorithm 5.1. Note that the transition from the edge map seen in Panel (a) to the one depicted in Panel (b) was achieved by a variation in the threshold value of the order of less than one tenth of a percent. In contrast the result obtained by Algorithm 5.1 retains its appearance robustly with respect to β and δ. Further, in numerous tests Canny's detector was not able to recognize the left edge of the square. However, note carefully that the left edge of the square is almost entirely visible in the edge map computed by Algorithm 5.1. Preliminary results suggest that including second order information of the smooth image I_s in the computation of the fuzzy edge map $\tilde{E}$ may further improve results in the presence of such edges. For instance it could be observed that extending Algorithm 5.1 in this manner allowed for the detection of the entire left edge of the particular image discussed here. A thorough investigation of such an extension may be an interesting topic for future work.

Finally we evaluate χ with respect to a K-Means segmentation. We want to obtain the best possible edges for the corrupted artificial image considered. In other words we want to discriminate back- and foreground of the image. Thus a natural choice of desired segments in the K-Means algorithm is $K = 2$. Figure 2.7 (h) and (i) shows the computed K-Means segments for the noisy artificial image. As briefly addressed in Section 2.3 K-Means clustering is obviously very prone to noise. A closer look at the underlying iterative strategy of the K-Means approach reveals the cause of its poor performance
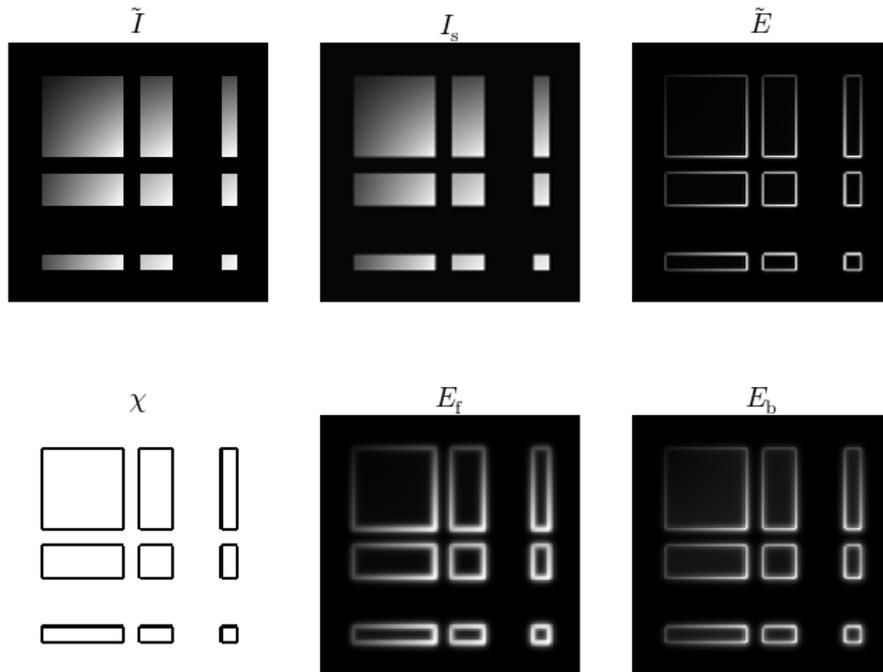
Figure 6.6: Result of Algorithm 5.1 for an artificial image using $\chi = 0$ as initialization and the parameter values $\beta = 5$, $\delta = 0.2$ and $\vartheta = 1$.

straightaway. The mean value computations in the Update Step are very sensitive to outliers, i.e., pixels of differing intensity. Thus the computed means are inaccurate which deteriorates the segments calculated in the Assignment Step. In addition, the lack of any penalization with respect to regularity of segments leads to excessive fragmentation (compare Panel (h) of the same Figure). As mentioned above, results can be improved though by using preprocessing techniques. However, K-Means clustering fundamentally relies on the assumption of piecewise constancy of the raw image. Thus, even in the noise free case a satisfactory segmentation could not be obtained (compare Panels (b) and (e) of the same Figure).

Note however, that K-Means clustering plays an important role in this work as it used to compute an initial guess for χ in Algorithm 5.1. Thus the next Section is devoted to analyzing the behavior of the developed approach with respect to different initializations for the binary edge map.

6.3 Initializing the Binary Edge Map

For the sake of simplicity we focus again on the noise free artificial image shown in Figure 2.2. For all simulations we used $\text{tol}_{\text{in}} = \text{tol}_{\text{out}} = 1.0e - 2$, $k_{\text{out}}^{\max} = 5$ and $k_{\text{in}}^{\max} = 20$ unless explicitly stated otherwise. We start by discussing the case of using no initial guess

for χ at all, i.e., we set $\chi = 0$ at the beginning of Algorithm 5.1. Note that $\delta > 0$ guarantees that the variational problems (5.4), (5.11) and (5.13) are still well posed. Figure 6.6 shows the computed binary edge map χ. Note that for the sake of better comparability the same parameter values as for the result shown in Figure 3.5 have been used. Obviously the difference to χ computed using K-Means clustering as initial guess is minuscule. Thus one may be tempted not to rely on any initial guess for χ at all. However, numerous tests showed that especially for more realistic images K-Means initialization can improve convergence of Algorithm 5.1. For instance the clean artificial image discussed here required one additional step in the very first run of the Inner Iteration when using $\chi = 0$ as initial guess. Moreover, under certain circumstances the initialization $\chi = 0$ may bring about the need to adapt the edge thickness parameter ϑ. For instance, a large value of β in connection with a too small ϑ may give $\chi = 0$ in the first step of the Inner iteration which causes trivial convergence. Hence K-Means clustering is employed as default initialization. However, we want to emphasize that convergence of Algorithm 5.1 does not rely upon it. Another natural choice for initializing χ is using a Canny edge map. Starting Algorithm 5.1 with the Canny edges depicted in Panels (b) and (c) of Figure 2.4 yields results that are visually identical to the one depicted in Figure 3.5. Compared to the K-Means initialization one additional step in the first run of the Outer Iteration is required for the edge map corresponding to $\tau = 0.04$. For $\tau = 0.05$ the Canny edges are already of very high quality and thus do not need much refinement. Nevertheless, this result confirms the robustness of Algorithm 5.1. A good start accelerates convergence but leads to (visually) the same result. However, the severe drawback of Canny's edge detector remains. The threshold τ has to be carefully calibrated for each image. In contrast, K-Means clustering may yield poorer results than a fine tuned Canny edge map. Nonetheless, the K-Means approach requires much less maintenance. In fact, using $K = 2$ provides acceptable initializations in most cases whereas Canny edge maps tend to severely deteriorate if τ is chosen inappropriately. Since a poor Canny edge map may not prevent convergence but potentially slows down Algorithm 5.1 we decided to initialize χ by employing K-Means clustering.

Note that not only χ may be initialized in various ways. It is of course also possible to employ different strategies for computing I_{s} and thus $\tilde{E}$. Recall our investigation of the Ambrosio–Tortorelli approach with respect to results obtained by Algorithm 5.1. We compared the Ambrosio–Tortorelli phase function ψ_ε with the computed fuzzy edge map $\tilde{E}$. Thus a natural question is: what happens if we set $\tilde{E} = 1 - \psi_\varepsilon$ in Algorithm 5.1? Thus we set $k_{\mathrm{out}}^{\max} = 1$ since $\tilde{E}$ is not updated using χ computed in the Inner Iteration. Further, we start with $\chi = 0$ so that the behavior of the Inner Iteration can be observed without potential interference of a particularly good or bad initialization. Note that the quality of such a mixed strategy crucially relies on suitable parameter values for J_{AT} given by (2.10). If α, κ and ε are chosen appropriately then χ is visually identical to the edge map seen in Figure 3.5. However, very weak edges in ψ_ε can make the fuzzy edge map $\tilde{E}$ rather "flat" in some regions. This causes the back- and foreground image to be almost identical in areas around weak edges which may impair the computation of χ in the Inner Iteration. Thus the benefit of such an approach is indistinguishable at the moment; not only because of the introduction of three new parameters via J_{AT} but also due to the much higher computational effort required to minimize J_{AT} in contrast to solving (4.63). Note that setting $I_{\mathrm{s}} = I_{\mathrm{AT}}$ and $\tilde{E} = \nabla I_{\mathrm{s}} = \nabla I_{\mathrm{AT}}$ yields visually identical results but obviously has the same drawbacks.

(a)	(b)
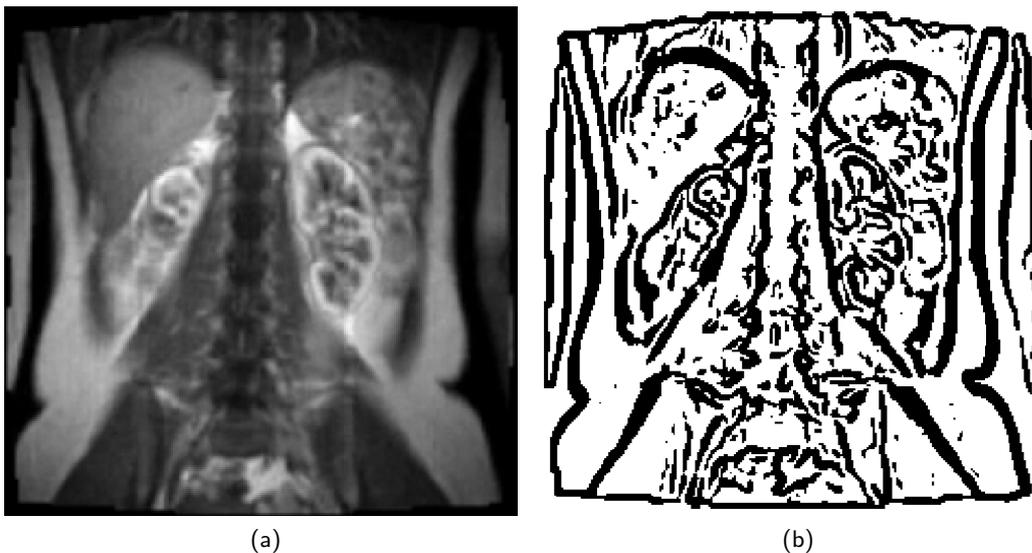
Figure 6.7: An edge map of a magnetic resonance image. Shown is the MR-image (a) and χ computed by Algorithm 5.1 using $\beta = 50$, $\delta = 0.8$ and $\vartheta = 1$ (b).

Another idea that has been tested is to use a TV-denoised approximation I_{TV} in Algorithm 5.1. In detail, we set $I_{\mathrm{s}} = I_{\mathrm{TV}}$ and $\tilde{E} = \nabla I_{\mathrm{TV}}$. For piecewise constant images this approach provides a valid initial guess for the first binary edge map computed in the Inner Iteration. However, in the case of images that manifest smooth intensity variations, like the artificial image discussed here, I_{TV} typically suffers from staircasing artifacts. Since pronounced staircasing is reflected in $\tilde{E}$ as well, convergence of the Inner Iteration may be impaired. Furthermore, TV denoising introduces a new parameter ν that has to be reasonably adjusted for each image. In contrast, I_{s} depends on the same parameters β and δ that are also used in the Inner Iteration. Moreover, the smooth image I_{s} was explicitly designed to overcome the assumption of piecewise constancy of the raw data.

In general it is of course possible to refine results obtained by Algorithm 5.1 by applying adequate preprocessing strategies to the raw image $\tilde{I}$. However, it is beyond the scope of this work to present a detailed review of modern image enhancement and restoration strategies. We refer the interested reader to [CS05] or [AK06].

6.4 Biomedical Applications

Having reviewed the behavior of Algorithm 5.1 using an artificial image we now focus on the applications that originally motivated the specific design of the presented approach. For all results in this section the default settings $\mathrm{tol_{in}} = \mathrm{tol_{out}} = 1.0e - 2$, $k_{\mathrm{out}}^{\max} = 5$ and $k_{\mathrm{in}}^{\max} = 20$ have been used. We start by considering a member of a sequence of dynamic contrast enhanced magnetic resonance images (DCE-MRIs) of a human torso as depicted in Panel (a) of Figure 6.7. In contrast to the artificial data used so far the MR-image manifests many small intertwined structures and fine textures. Moreover, the image is clearly not piecewise constant and exhibits large local intensity variations due to the presence of contrast agent. Panel (b) of the same Figure shows the edge map that was computed by Algorithm 5.1 in two Outer Iterations (comprising of first three

(a)   (b)
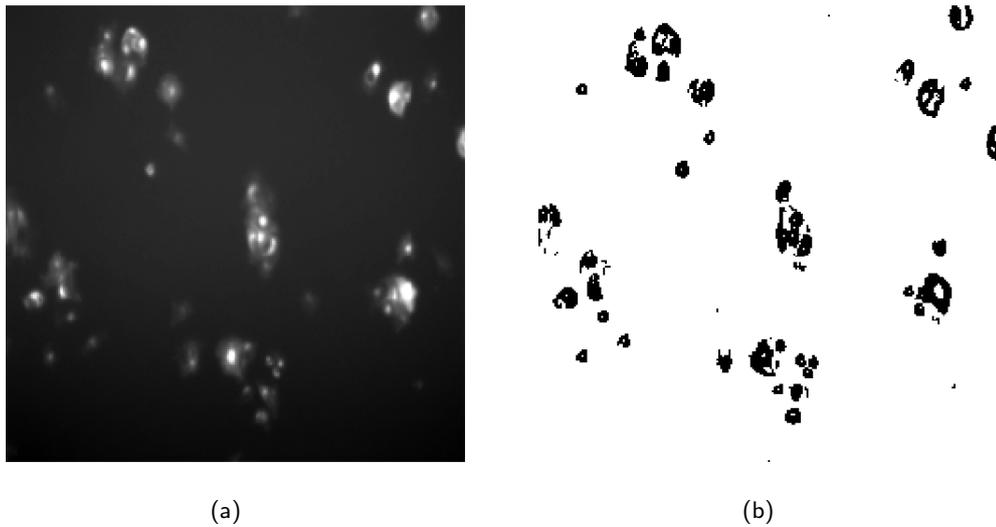
Figure 6.8: An edge map of a microscopic image. Shown is the raw photomicrograph (a) and χ
computed by Algorithm 5.1 using $\beta = 5e3$, $\delta = 0.01$ and $\vartheta = 1$ (b).

then one Inner Iterations to reach convergence). By using slightly increased values for
β and δ the presented approach is capable of outlining major objects like the liver while
not cluttering the edge map with many small details. Simultaneously Algorithm 5.1 is
sensitive enough to capture fine details like physiological structures inside the kidneys
that are highlighted by contrast agent. By comparison, Figure 2.13 shows an Ambrosio–
Tortorelli phase function for the same image before and after thresholding. Observe that
edges that are particularly pronounced in χ appear as multi-layered lines in the thresholded
phase function. Moreover, organ contours as well as smaller details like the interior of the
kidneys are hardly distinguishable. Note that the thresholding result might be improved
by exhaustive parameter testing since slight variations in the threshold cause significant
changes in the edge map. Similarly, a Canny edge map may yield satisfactory results but
requires extensive fine tuning of the threshold. Conversely, Algorithm 5.1 proves again its
stability with respect to β and δ; varying β by ± 10 and δ by ± 0.1 has no visible influence
on the result.

To illustrate the role of the edge thickness parameter ϑ in practice consider Panel (a) of
Figure 6.8. Shown is a microscopic image of cancer cells in a petri dish. From an image
processing point of view this photomicrograph poses similar challenges as the MR-image
discussed above. A chemical agent causes cancerous cells to illuminate while healthy cells
show no or little response. Thus the image also exhibits large local intensity variations
and consists of many fine grained structures. An additional challenge is the presence of
background noise. Though the image looks rather clean it is in fact corrupted by noise
that is roughly of the order of poorly illuminated cells. This makes a clear separation
of cell nuclei from the background difficult. To account for the noise corruption we thus
significantly increased β such that I_s and hence $\tilde{E}$ become sufficiently smooth. However,
to avoid losing fine structures within cells we simultaneously decreased δ. This lowers
the value of $\beta\delta$ and hence avoids oversmoothing on edges (i.e., where $\chi = 0$) such that
intracellular details are preserved. Panel (b) of Figure 6.8 shows the computed edge map
for $\vartheta = 1$. Note that indeed cell nuclei are outlined by pronounced edges and thus clearly

(a)　　　　　　　　　　　　　　　　　　　(b)
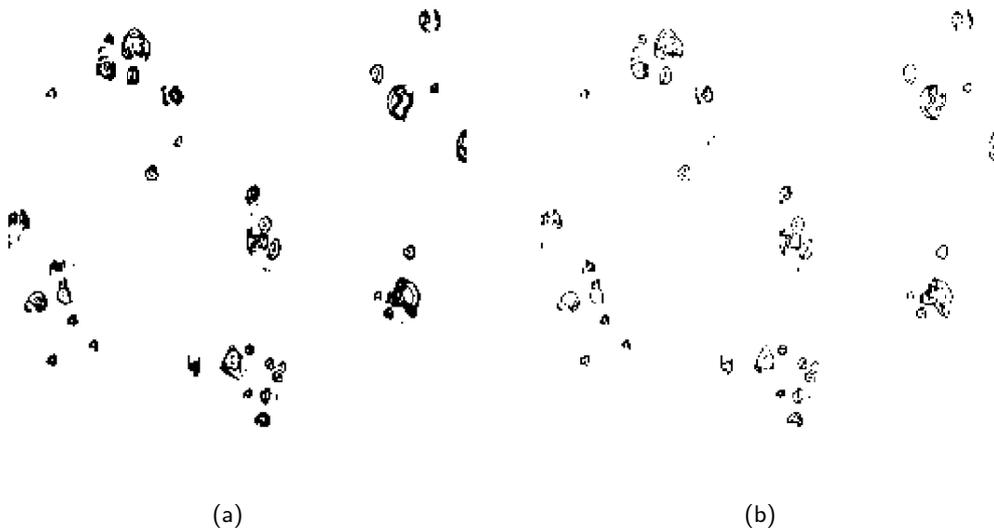
Figure 6.9: Illustration of the role of ϑ. Shown is χ computed by Algorithm 5.1 for the microscopic image depicted in Panel (a) of Figure 6.8 using $\vartheta = 0.5$ (a) and $\vartheta = 0.25$ (b) with $\beta = 5e3$ and $\delta = 0.01$.

visible in χ. In contrast the edge maps seen in Figure 6.9 computed using $\vartheta = 0.5$ and $\vartheta = 0.25$ exhibit considerably finer lines. Depending on the requirements of the specific application Algorithm 5.1 is thus capable of generating edges of various strengths. Note carefully, however, that the particular topological structure of the computed edges is *not* altered by the decrease in ϑ. In fact, for $\vartheta \to 0$ the computed edge maps show a behavior analogous to Ambrosio–Tortorelli phase functions as $\varepsilon \to 0$. It is not within the scope of this work to perform a rigorous mathematical investigation of this limiting phenomenon but an asymptotic analysis of the presented approach for $\vartheta \to 0$ may be rewarding.

The next chapter is devoted to a brief introduction to image registration. It will be demonstrated that the developed approach for the computation of binary edge maps lends itself handily to be embedded in an iterative registration strategy specifically tailored for image sequences that prove to be challenging for classic registration techniques.

The Image Registration Problem

A common problem in image processing is to combine data from various sources to generate an image or to enhance an existing image's quality. In astronomy, for instance, one often wants to assemble multiple color channels and/or wavelengths to visualize measurements obtained by space telescopes. A common task in medical imaging consists of comparing patient data before and after treatment. State of the art surgery planing involves incorporating data from different sources, e.g., magnetic resonance and computerized tomography imaging [FHM00, Chap. 8]. All these applications require the computation of an explicit coordinate transform between points of one image and their corresponding counterparts in the other. This task is called *image registration*.

The methods used to solve registration problems are as diverse as the underlying applications. A possible classification may be imposed by categorizing techniques based on the kind of transformation that is computed and on the notion of image similarity used. An exhaustive review of modern techniques in image registration is given in [Mod04]. In the following we will focus on registration approaches used in the context of this work.

The objective of this chapter is to develop a strategy for the robust and efficient registration of edges. More specifically, we want to embed the approach introduced in Chapter 3 in an image registration framework. Of particular interest will be sequential image data exhibiting pronounced intensity variations, e.g. magnetic resonance images in the presence of contrast agent, since these settings prove to be challenging for standard intensity based registration schemes. Nevertheless, the development of the registration strategy presented below was driven by an application coming from biophysics [FKPP11].

7.1 Image Similarity

We start by introducing terminology commonly used in the context of image registration. Let again $\Omega := (0,1)^2$ denote the image domain and let I_0 and I_1 be two images on Ω. The simplest way to formulate a registration problem is to find a displacement field $\boldsymbol{w} : \mathbb{R}^2 \to \mathbb{R}^2$ such that $I_0(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})) \approx I_1(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$. In this context we refer to I_0 as *the template* or *the moving image* whereas I_1 is the so-called *reference* or *fixed image*. Thus, in words, the task is to find a vector field $\boldsymbol{w}$ that deforms the template I_0 in such a way that $I_0 \circ (\mathrm{id} + \boldsymbol{w})$ is *close* to the reference I_1. Obviously, we have to clarify what we mean by "close" in this context, i.e., we have to specify an appropriate *similarity measure*. The choice of a suitable notion of image similarity depends solely upon the registration problem at hand. Recall that the objective was to register edge maps. Thus let χ_0 and χ_1 denote the template and reference edge maps, respectively, and define $\Gamma_i := \mathrm{supp}(\chi_i)$ for $i = 0, 1$. Thus Γ_0 and Γ_1 are two-dimensional curves in Ω. Note that in contrast to previous chapters $\chi_i = 1$ *only* on Γ_i whereas before $\chi = 1$ everywhere excluding edges. We assume that both curves have finite length, i.e., the Hausdorff measure $\mathcal{H}(\Gamma_i)$ is finite. While this

(a)   (b)   (c)
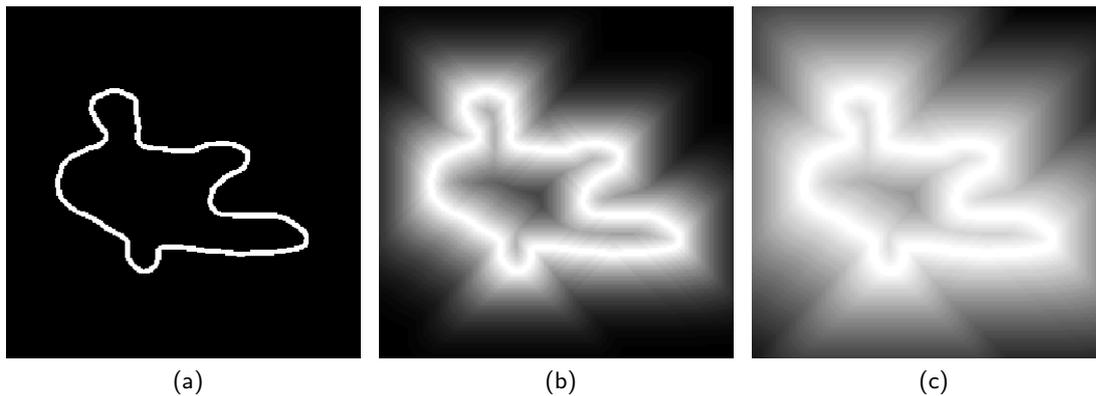
Figure 7.1: The effect of ε in (7.3). Shown is an artificial edge map (a) and its blurred versions according to (7.3) for progressively larger values of ε (b) and (c).

may look like a simplification, $\mathcal{H}(\Gamma_i) < \infty$ introduces the concept of zero measure sets in a registration setting. This has a fundamental impact on choosing an applicable similarity measure. Consider for instance the *sum of squared intensity differences* (SSID) which is one of the most widely used similarity measures [FHM00, Chap. 8]. In our notation it takes the form

$$(7.1) \qquad \frac{1}{2} \int_\Omega |\chi_0 \circ (\mathrm{id} + \boldsymbol{w}) - \chi_1|^2 \, d\boldsymbol{x},$$

for which we make the following crucial observation. Since $\Gamma_i := \mathrm{supp}(\chi_i)$ and $\mathcal{H}(\Gamma_i) < \infty$ the sets Γ_i have measure zero $|\Gamma_i| = 0$. Thus the trivial deformation $\boldsymbol{w} = 0$ minimizes (7.1) for any χ_0 and χ_1. This renders the SSID measure (7.1) useless in this context. Thus one may consider an approach tailored specifically to quantify difference of edges such as the *Hausdorff distance*. Hence consider

$$(7.2) \qquad d_{\mathrm{H}}(\Gamma_0, \Gamma_1) := \max \left(\sup_{\boldsymbol{x} \in \Gamma_0} d_{\Gamma_1}(\boldsymbol{x}), \sup_{\boldsymbol{x} \in \Gamma_1} d_{\Gamma_0}(\boldsymbol{x}) \right),$$

with $d_{\Gamma_i}(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in \Gamma_i} |\boldsymbol{x} - \boldsymbol{y}|$. Usage of the Hausdorff distance has a long tradition in computer graphics and image processing in general [HKR93]. It is commonly employed in the context of shape recognition tasks, e.g., face detection [JKF01]. Nevertheless, the Hausdorff distance has also been successfully applied in image registration. However, the downside of using (7.2) is its computational complexity (see, e.g., [FJSY09]) which is sometimes counteracted by imposing explicit restrictions on the computed deformation field [KKS09]. On the other hand, extended formulations such as the one presented in [DR06] require sufficient regularity of the problem which cannot be guaranteed here.
The approach developed in [FKPP10] was guided by the idea of enhancing the SSID measure (7.1) in a way that mimics some intrinsic features of (7.2) to obtain a similarity measure that is not only applicable to edge maps but also computationally efficient. Thus we adopt the idea of Ambrosio–Tortorelli phase functions discussed in Section 2.5 and

define

$$(7.3) \qquad \chi_i^\varepsilon(\boldsymbol{x}) := \begin{cases} 1 - d_{\Gamma_i}(\boldsymbol{x})/\varepsilon, & \text{if } d_{\Gamma_i}(\boldsymbol{x}) \le \varepsilon, \\ 0, & \text{otherwise,} \end{cases} \text{ for } i = 0, 1,$$

where $\varepsilon > 0$ is a *blurring parameter*. Note the similarity of this definition to (2.7). Hence χ_i^ε is one on Γ_i and smoothly decreases to zero off Γ_i. Thus χ_i^ε is supported on a band of width 2ε around Γ_i. That means (7.3) extends the support of χ_i depending on the magnitude of ε. Observe carefully that now $|\text{supp}(\chi_i^\varepsilon)| > 0$, which thus permits the use of the SSID measure. Hence we introduce the following distance measure for blurred edge maps

$$(7.4) \qquad S^\varepsilon[\boldsymbol{w}] := \frac{1}{2} \int_\Omega |\chi_0^\varepsilon \circ (\text{id} + \boldsymbol{w}) - \chi_1^\varepsilon|^2 \, d\boldsymbol{x}.$$

Several important observations should be made at this point. First, the term "blurring parameter" for ε is indeed justified as depicted by Figure 7.1 which illustrates the effect of (7.3). Note that the technique used to obtain χ_i^ε can be seen as a distance transform which is a well known tool in image processing (see, e.g., [Jai89]). Several authors have used distance transforms in the context of image registration, e.g., [HB06] employed constrained distances whereas [PRR02] developed a variational approach to match distance functions. Note further that the striking similarity between (7.3) and (7.2) is the distance function d_{Γ_i}. Thus (7.3), just like (7.2), is potentially expensive to compute. Hence in practice (7.3) is calculated via a marching scheme based on successive discrete convolutions. Section 8.2 addresses in further detail the relation between the Hausdorff distance and S^ε.

7.2 Elastic Deformations

In addition to determining a suitable similarity measure, choosing an appropriate class of feasible deformations has a fundamental impact on the performance of the registration. Based on this choice one either refers to *parametric* or *non-parametric image registration* [Mod04]. In parametric registration explicit restrictions on the desired deformation are imposed. Thus for instance one is only interested in affine linear transformations, i.e., one looks for deformations fields $\boldsymbol{w}$ of the form $\boldsymbol{w} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with some matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ satisfying $\det(A) > 0$ and $\boldsymbol{b} \in \mathbb{R}^2$. A special case of affine linear transformations are so-called *rigid transformations* satisfying $\boldsymbol{w} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}$ where $\boldsymbol{W} \in \mathbb{R}^{2\times 2}$ is an orthogonal matrix. The name rigid refers to the fact that these transformations are solely composed of rotations and translations. Note that non-parametric techniques are usually computationally very efficient; however, narrowing down the degrees of freedom for $\boldsymbol{w}$ may impair the quality of the registration significantly. In the case of an affine linear registration, for instance, every point $\boldsymbol{x}$ in the image is registered by the same mapping rule $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$. This makes it impossible to align only small local features while leaving the rest of the images unchanged.

Thus deformation fields obtained by a non-parametric registration approach will be of primary interest in the following. We will utilize techniques from continuum mechanics to deduce an energy functional that is minimized by a certain class of deformations. The subsequent derivation follows the presentation given in [KR05]. We introduce
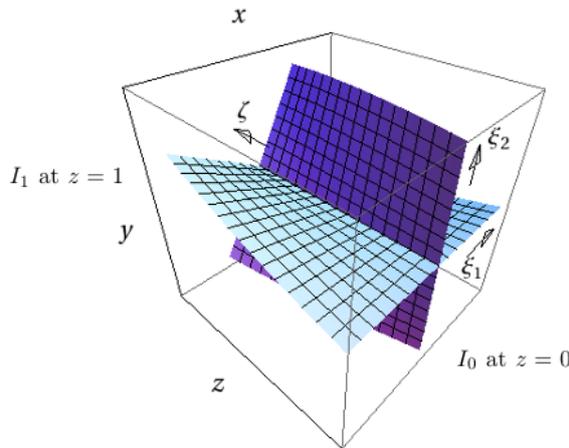
Figure 7.2: Sketch of the image cube Q illustrating the continuum mechanical setup.

the "image cube" $Q := \Omega \times (0,1) \subset \mathbb{R}^3$ thus

$$Q = \left\{ (x,y,z) \in \mathbb{R}^3 \,|\, 0 < x, y, z < 1 \right\}.$$

Returning to the notation from above let I_0 and I_1 denote again the template and reference image respectively. Suppose I_0 and I_1 are placed on the front and back face of Q denoted by Ω_0 and Ω_1 in this context. Figure 7.2 shows a sketch. We use the surfaces depicted in said Figure to introduce an alternative coordinate system. Let $\boldsymbol{\xi} := (\xi_1, \xi_2)$ be curvilinear coordinates such that all but one component are constant along the shown surfaces. The intersection of the surfaces forms a trajectory connecting a point in the template image I_0 to its counterpart in the reference image I_1. We parameterize such trajectories through Q according to $\zeta = z$. Hence (x, y, z) refers to a spatial or *Eulerian* coordinate system whereas (ξ_1, ξ_2, ζ) are *Lagrangian* coordinates (for an introduction to these concepts see, e.g., [Mes06]). With $\boldsymbol{x} = (x, y)$ we initialize $\boldsymbol{\xi}(\boldsymbol{x}, 0) = \boldsymbol{x}$ such that displacements in Q are given by $\boldsymbol{d}(\boldsymbol{x}, z) = \boldsymbol{x} - \boldsymbol{\xi}(\boldsymbol{x}, z)$. Consequently, trajectories originating in Ω_0 can be written as $\boldsymbol{x}(\boldsymbol{\xi}, \zeta)$. Note that in general not every point Ω_0 has a corresponding point in Ω_1. Thus we introduce

$$\Omega_0^c := \{ \boldsymbol{\xi} \in \Omega_0 \,|\, \boldsymbol{x}(\boldsymbol{\xi}, \zeta) \in Q, \,\forall \zeta \in (0,1) \},$$

to denote the subset of Ω_0 that consists of points that find like counterparts in Ω_1.

We make the assumption that the reference image is the result of an elastic deformation acting on the template image. In continuum mechanics a very popular tool to characterize the strain occurring in a body under an elastic deformation is the so-called *right Cauchy–Green strain tensor* [Cia88]

$$\boldsymbol{C}(\zeta) := \nabla_{\boldsymbol{\xi}} \boldsymbol{x}^\top \nabla_{\boldsymbol{\xi}} \boldsymbol{x}.$$

Note that *deforming* a body means to change distances between particle pairs in the body and thus to alter the body's form. A rigid body motion, on the other hand, induces a uniform displacement of all particles albeit *not* affecting the particles' mutual distances. This is reflected by the fact that $\boldsymbol{C}(\zeta) = \boldsymbol{I}$ for rigid body motions, where $\boldsymbol{I}$ denotes the identity matrix. Thus the *Green–St.Venant strain* $\boldsymbol{E} := \frac{1}{2}(\boldsymbol{C} - \boldsymbol{I})$ quantifies the deviation of a deformation from rigidity. Using this strain, an expression for the elastic potential

energy in a body can be derived [Cia88]

$$(7.5) \qquad W[\boldsymbol{E}] = \int_{\Omega_0^c} \lambda \mathrm{tr}(\boldsymbol{E})^2 + 2\mu \left|\boldsymbol{E}\right|^2 d\boldsymbol{\xi},$$

where $\mathrm{tr}(\boldsymbol{E})$ denotes the trace of $\boldsymbol{E}$ and $\lambda, \mu > 0$ are the so-called *Navier–Lamé constants* quantifying a body's elastic properties. Sometimes μ is also called *shear modulus* and λ is referred to as *bulk modulus*. Note that the shear term $2\mu \left|\boldsymbol{E}\right|^2 = 2\boldsymbol{E} : \boldsymbol{E}$ makes $W(\boldsymbol{E})$ non-linear and thus potentially expensive to compute in practice. Thus using the introduced expression for internal displacements in Q we linearize the Green–St.Venant strain

$$2\boldsymbol{E} = \boldsymbol{C} - \boldsymbol{I} = \nabla_\xi \boldsymbol{d} + \nabla_\xi \boldsymbol{d}^\top + \nabla_\xi \boldsymbol{d}^\top \nabla_\xi \boldsymbol{d} \approx \nabla_\xi \boldsymbol{d} + \nabla_\xi \boldsymbol{d}^\top,$$

and thus obtain the following approximation

$$(7.6) \qquad W[\boldsymbol{E}] \approx \int_{\Omega_0^c} \lambda \left(\nabla_{\boldsymbol{\xi}} \cdot \boldsymbol{d}\right)^2 + \frac{\mu}{2} \left|\nabla_\xi \boldsymbol{d}^\top + \nabla_\xi \boldsymbol{d}\right|^2 d\boldsymbol{\xi}.$$

Hence we obtain the *linearized elastic potential* of $\boldsymbol{w}$

$$(7.7) \qquad P[\boldsymbol{w}] := \frac{1}{2} \int_\Omega \lambda \left(\nabla \cdot \boldsymbol{w}\right)^2 + \mu \left|\nabla \boldsymbol{w}^\top + \nabla \boldsymbol{w}\right|^2 d\boldsymbol{x}.$$

The procedure of minimizing a linear combination of some similarity measure and P is called *elastic registration* which will be the method of choice in this work.

Elastic registration has been used by many authors, e.g., [PSRS99] or [KR05], and has proven to be a reliable technique in medical imaging applications. Note carefully, however, that (7.6), as being a linearization of (7.5), can only be expected to approximate (7.5) for relatively modest deformations. Hence P implicitly relies on the assumption that $\boldsymbol{w}$ induces only small changes. Thus the use of elastic registration may not be adequate if template and reference image vary significantly. Nevertheless, an elastic penalizer has been used in [FKPP10] to register blurred edge maps arising from cuts through three dimensional heart models. Thus it could not always be guaranteed that deviations in the template and reference image were sufficiently small. However, the obtained results were of very high quality and the employed strategy proved to be quite robust and reliable. Hence the approach was also used outside the specific application for which it was originally designed.

7.3 An Optimization Problem

Having selected a similarity measure (7.1) and a desired class of deformations (7.7) we can now set up a cost functional

$$(7.8) \qquad \begin{aligned} J_{\mathrm{reg}}^\varepsilon[\boldsymbol{w}] :={}& S^\varepsilon[\boldsymbol{w}] + P[\boldsymbol{w}] \\ ={}& \frac{1}{2} \int_\Omega \left|\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) - \chi_1^\varepsilon\right|^2 + \lambda \left(\nabla \cdot \boldsymbol{w}\right)^2 + \mu \left|\nabla \boldsymbol{w}^\top + \nabla \boldsymbol{w}\right|^2 d\boldsymbol{x}, \end{aligned}$$
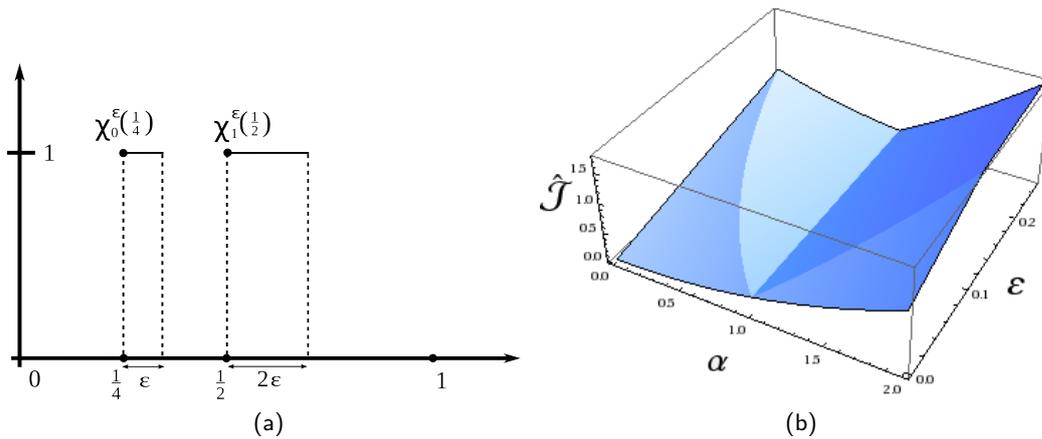
Figure 7.3: A one-dimensional registration problem. (a) The problem setup (b) Surface plot of $\hat{\mathcal{J}}(\varepsilon, \alpha)$ for $0 \leq \varepsilon \leq \frac{1}{4}$ and $0 \leq \alpha \leq 2$ with $\mu = \frac{1}{4}$.

for some fixed $\varepsilon > 0$. Thus elastically registering χ_0 to χ_1 means in our setting to solve the minimization problem

$$(7.9) \qquad \qquad \inf_{\boldsymbol{w}} J_{\text{reg}}^{\varepsilon}[\boldsymbol{w}].$$

Similar to the presentation given in Chapter 3 we start by discussing properties of (7.8) first and address mathematical details later in Chapter 8. Note that (7.9) is a stand alone minimization problem for every $\varepsilon > 0$. Thus (7.9) can also be seen as a family of optimization problems depending on ε. Hence besides discussing the existence of solutions to (7.9) for fixed ε an obvious question concerns the behavior of solutions as ε gets progressively smaller. The following one-dimensional considerations (following [FKPP10]) may shed some light on the intrinsic features of (7.9) as $\varepsilon \to 0$.

Assume for now that $\Omega = (0, 1) \subset \mathbb{R}$ and let

$$\chi_0(x) := \begin{cases} 1, & x = \frac{1}{4}, \\ 0, & \text{otherwise}, \end{cases} \quad \text{and} \quad \chi_1(x) := \begin{cases} 1, & x = \frac{1}{2}, \\ 0, & \text{otherwise}. \end{cases}$$

Similar to the two dimensional problem discussed above we want to elastically register χ_0 to χ_1. Thus we introduce the following cost functional

$$\hat{J}[w] := \hat{S}[w] + \mu \hat{P}[w],$$

with

$$\hat{S}[w] := \int_0^1 |\chi_0 \circ (\text{id} + w) - \chi_1|^2 \, dx,$$

and

$$\hat{P}[w] := \int_0^1 |w'(x)|^2 \, dx,$$

where μ is a regularization parameter (playing the role of both Navier–Lameé constants in one dimension). Note that also in this one-dimensional setting both χ_0 and χ_1 are only supported on sets of measure zero (on single points). Hence $\hat{S}[0] = 0$, meaning that

again the trivial deformation $w = 0$ is an (unwanted) minimizer. Thus following the above presented strategy we introduce

$$\chi_0^\varepsilon(x) := \begin{cases} 1, & \frac{1}{4} \leq x \leq \frac{1}{4} + \varepsilon, \\ 0, & \text{otherwise}, \end{cases} \quad \text{and} \quad \chi_1^\varepsilon(x) := \begin{cases} 1, & \frac{1}{2} \leq x \leq \frac{1}{2} + 2\varepsilon, \\ 0, & \text{otherwise}, \end{cases} \quad \text{for} \quad 0 < \varepsilon \leq \frac{1}{4},$$

in analogy to (7.3). Figure 7.3 depicts a sketch. We similarly update the cost functional to be

$$\hat{J}^\varepsilon[w] := \hat{S}^\varepsilon[w] + \mu \hat{P}[w],$$

with

$$\hat{S}^\varepsilon[w] := \int_0^1 |\chi_0^\varepsilon \circ (\mathrm{id} + w) - \chi_1^\varepsilon|^2 \, dx.$$

Due to the plain structure of this problem we can determine the form of wanted deformations a priori. Looking at Figure 7.3(a) the deformation we are looking for is obviously given by $w^\star(x) := x$. Thus we may narrow down the problem to minimizing $\hat{J}^\varepsilon$ under the constraint that $w_\alpha = \alpha x$ with $\alpha \in \mathbb{R}$ and some fixed $0 < \varepsilon \leq \frac{1}{4}$. For deformations of this form the penalizer may be computed explicitly

$$\hat{P}[w_\alpha] = \int_0^1 |\alpha|^2 \, dx = \alpha^2,$$

which immediately yields

$$(7.10) \qquad\qquad \hat{P}[0] = 0 \quad \text{and} \quad \hat{P}[w^\star] = 1.$$

We can analogously derive an explicit expression for the similarity measure at $w = 0$. Thus suppose α satisfies

$$(7.11) \qquad\qquad (1 + \alpha)(\frac{1}{4} + \varepsilon) \leq \frac{1}{2},$$

which means the support of $\chi_0^\varepsilon \circ (\mathrm{id} + w_\alpha)$ remains to the left of the support of χ_1^ε. Then

$$\hat{S}^\varepsilon[w_\alpha] = \int_0^1 |\chi_0^\varepsilon(x + \alpha x) - \chi_1^\varepsilon(x)|^2 \, dx$$
$$= (1 + \alpha) \left(\left(\frac{1}{4} + \varepsilon - \frac{1}{4} \right) + \left(\frac{1}{2} + 2\varepsilon - \frac{1}{2} \right) \right)$$
$$= (\alpha + 3)\varepsilon.$$

Since $0 < \varepsilon \leq \frac{1}{4}$, (7.11) holds for $\alpha = 0$ which thus implies

$$(7.12) \qquad\qquad \hat{S}^\varepsilon[0] = 3\varepsilon.$$

On the other hand, for $w^\star$ we have $\alpha = 1$. Hence suppose $\frac{1}{2} < (1 + \alpha)(\frac{1}{4} + \varepsilon) \leq \frac{1}{2} + 2\varepsilon$, i.e., the support of $\chi_0^\varepsilon \circ (\mathrm{id} + w_\alpha)$ remains inside the support of χ_1^ε. Then

$$\hat{S}^\varepsilon[w_\alpha] = \left(\frac{1}{2} - (1 + \alpha)\frac{1}{4} \right) + \left(\frac{1}{2} + 2\varepsilon - (1 + \alpha)(\frac{1}{4} + \varepsilon) \right) = \frac{1}{2}(1 - \alpha)(1 + 2\varepsilon),$$

which implies for $\alpha = 1$ that

(7.13)
$$\hat{S}^\varepsilon[w^\star] = 0.$$

Thus for

(7.14)
$$\varepsilon < \min\left\{\frac{1}{4}, \frac{\mu}{3}\right\},$$

combining (7.10), (7.12) and (7.13) gives

(7.15)
$$\begin{aligned}
\hat{J}^\varepsilon[0] = \hat{S}^\varepsilon[0] + \mu\hat{P}[0] &= 3\varepsilon + \mu \cdot 0 \\
&< 3\frac{\mu}{3} \\
&= \hat{S}^\varepsilon[w^\star] + \mu\hat{P}[w^\star] = \hat{J}^\varepsilon[w^\star],
\end{aligned}$$

which means that the trivial deformation yields a lower cost than the wanted displacement *despite* the fact that $\varepsilon > 0$. This observation is of fundamental importance for the development of a robust registration strategy since it is not a peculiarity of the one-dimensional setting. Consequently, this phenomenon motivated the design of an iterative registration approach that avoids the computation of trivial deformations as ε gets smaller. We explain the developed method by thoroughly investigating the presented example.

What led to the unwanted result above was the special choice of ε in (7.14). Inequality (7.15) only holds since ε was small enough compared to the regularization parameter μ. Thus the "right" choice of ε depends on the value of the regularization parameters. To illustrate this concept let $\hat{\mathcal{J}}(\varepsilon, \alpha) := \hat{J}^\varepsilon(\alpha x)$ with $\mu = \frac{1}{4}$. Figure 7.3 shows $\hat{\mathcal{J}}(\varepsilon, \alpha)$ for $\varepsilon \in [0, \frac{1}{4}]$ and $\alpha \in [0, 2]$. Note that $\hat{J}^\varepsilon$ is minimized by the wanted deformation $w^\star$ for sufficiently large values of ε while $\alpha = 0$ and thus the trivial deformation is a minimizer for small ε. Hence choosing ε large may seem like a possible remedy to this problem. However, large values of ε induce heavy blurring in the edge maps (compare Figure 7.1) and thus may "wash out" local details in the curves Γ_i thereby possibly impairing the outcome of the registration. However, a deformation computed for such a large ε may still be a valid initial guess for a registration. Thus we start by computing the global minimizer of $\hat{J}^\varepsilon$ for a sufficiently large ε. We use the calculated "rough" deformation field as initialization for the registration with a smaller ε. The updated deformation field is used as initial guess for the registration with an even smaller ε and so forth. Since the very first deformation was computed for some large ε, this iterative procedure guarantees that solutions do not become trivial as $\varepsilon \to 0$. In the language of optimization we start by looking for the global minimum of $\hat{J}^\varepsilon$ but favor local minimizers as ε gets smaller. Algorithm 9.1 presents a comprehensive pseudo-code that illustrates this strategy and addresses practical details like stopping criteria, the discretization used and specifics of the implementation.

The next chapter is devoted to the mathematical analysis of the presented approach and sets up the optimization framework that is used in practice.

Analysis of The Registration Approach

We concluded the previous chapter by studying a one-dimensional problem that served as an object lesson for sketching an iterative registration strategy for edge maps. In the following we present a more rigorous discussion of the minimization problems which illustrates from an analytical point of view the reasoning behind the developed approach. We establish a common ground for the discussion below by making the following supposition.

Assumption 9. *Let* $\Omega := (0,1)^2$, $\chi_0^\varepsilon \in W^{1,\infty}(\mathbb{R}^2)$ *and* $\chi_1^\varepsilon \in L^\infty(\Omega)$ *with* $\varepsilon > 0$ *fixed. Further, let* $\lambda > 0$ *and* $\mu > 0$.

Note that the condition on χ_0^ε may seem like a strong regularity requirement. However, provided the curve Γ_0 is smooth enough, χ_0^ε satisfies Assumption 9 automatically. The distance functions d_{Γ_i} appearing in the Definition (7.3) of χ_i^ε are Lipschitz continuous with Lipschitz constant one [Zaj83, Sec. 3]. If Γ_0 is sufficiently regular then χ_0^ε has an essentially bounded weak derivative a.e. in Ω which is inherited from d_{Γ_0}. This is a consequence of Rademacher's Theorem [Eva08, Sec. 5.8, Th. 6] that states that a Lipschitz function is differentiable almost everywhere. Then by extending χ_0^ε by zero outside of Ω we get $\chi_0^\varepsilon \in W^{1,\infty}(\mathbb{R}^2)$. Note further that this implies that χ_0^ε is Lipschitz continuous on $\mathbb{R}^2$ (see Part 1 of the proof of Theorem 4 in [Eva08, Sec. 5.8.2.b]).

The results stated below have been presented first in [FKPP10]. Thus we only provide the general idea of argumentation and refer the reader to [FKPP10] for details. Similarly, Section 8.2 addresses the relation between S^ε and the Hausdorff distance (7.2) by discussing results previously published in [CFK04].

8.1 Existence of Solutions to the Registration Problem

We start by stating (7.9) more precisely. With a view to $J_{\text{reg}}^\varepsilon$ given by (7.8) the natural space for minimizing this functional is $H^1(\Omega)$. Thus we rewrite (7.9) in more detail

$$(8.1) \qquad \min_{\boldsymbol{w} \in H^1(\Omega)} J_{\text{reg}}^\varepsilon[\boldsymbol{w}], \quad \text{for} \quad \varepsilon > 0.$$

First we want to guarantee that this problem has a solution, i.e., the use of min instead of inf is indeed justified. A standard way to prove existence of a solution to an infinite dimensional minimization problem is the so-called *direct method of the calculus of variations* (see for instance [AK06, Sec. 2.1]). In the present context it may be summarized as follows. Since $J_{\text{reg}}^\varepsilon[\boldsymbol{w}] > 0$ for every $\boldsymbol{w} \in H^1(\Omega)$ and there exists some $\boldsymbol{w}$ such that $J_{\text{reg}}^\varepsilon[\boldsymbol{w}] < \infty$, e.g., $\boldsymbol{w} = 0$, the infimum $\inf J_{\text{reg}}^\varepsilon[\boldsymbol{w}]$ on $H^1(\Omega)$ is finite. In other words $J_{\text{reg}}^\varepsilon$ is bounded from below. Thus there exists a minimizing sequence $\{\boldsymbol{w}_k\}_{k \geq 1} \subset H^1(\Omega)$ such

that

$$(8.2) \qquad \lim_{k \to \infty} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}_k] = \inf_{\boldsymbol{w} \in H^1(\Omega)} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}].$$

Assume $J_{\mathrm{reg}}^{\varepsilon}$ is coercive on $H^1(\Omega)$, i.e.,

$$J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}] \geq \alpha \, \|\boldsymbol{w}\|_{H^1(\Omega)}^2, \quad \forall \boldsymbol{w} \in H^1(\Omega),$$

for some $\alpha > 0$ then $\{\boldsymbol{w}_k\}_{k \geq 1}$ is bounded in $H^1(\Omega)$. Since $H^1(\Omega)$ is reflexive this implies the existence of a subsequence $\{\boldsymbol{w}_{k_l}\}_{l \geq 1}$ of $\{\boldsymbol{w}_k\}_{k \geq 1}$ that converges weakly in $H^1(\Omega)$ to some $\boldsymbol{w}^* \in H^1(\Omega)$ (compare Appendix B.2). Suppose further that $J_{\mathrm{reg}}^{\varepsilon}$ is lower semicontinuous, i.e., $\liminf_{k \to \infty} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{z}_k] \geq J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{z}^*]$ for any sequence $\{\boldsymbol{z}_k\}_{k \geq 1} \subset H^1(\Omega)$ converging (weakly or strongly) to $\boldsymbol{z}^*$. Then (8.2) yields

$$\inf_{\boldsymbol{w} \in H^1(\Omega)} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}] = \lim_{k \to \infty} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}_k] = \lim_{l \to \infty} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}_{k_l}] \geq J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}^*] \geq \inf_{\boldsymbol{w} \in H^1(\Omega)} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}],$$

or equivalently $J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}^*] = \min_{\boldsymbol{w} \in H^1(\Omega)} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{w}]$. Thus if we can show coercivity and lower semicontinuity of $J_{\mathrm{reg}}^{\varepsilon}$, existence of a solution to (8.1) is readily established. However, the following considerations show that $J_{\mathrm{reg}}^{\varepsilon}$ is *not* coercive on $H^1(\Omega)$. Let

$$\mathrm{RM}(\Omega) := \left\{ \boldsymbol{w} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{c} \,\middle|\, \boldsymbol{W} \in \mathbb{S}^2, \boldsymbol{c} \in \mathbb{R}^2, \boldsymbol{x} \in \Omega \right\},$$

denote the space of *infinitesimal rigid motions* where

$$\mathbb{S}^2 := \left\{ \boldsymbol{W} \in \mathbb{R}^{2 \times 2} \,\middle|\, \boldsymbol{W} + \boldsymbol{W}^{\top} = 0 \right\},$$

is the space of skew symmetric matrices. Note that $\boldsymbol{W}$ is orthogonal for rigid motions whereas $\boldsymbol{W} \in \mathbb{S}^2$ for *infinitesimal* rigid motions. This comes from the fact that every orthogonal matrix can be written as matrix exponential of a skew symmetric matrix (compare, e.g., [GX02] and references therein) which is used in the theory of infinitesimal rotations (for an overview see, e.g., [Mes06, Sec. 2.1.2]). For $\boldsymbol{u} \in \mathrm{RM}(\Omega)$ we obviously have $\nabla \boldsymbol{u} = \boldsymbol{W}$ and thus $P[\boldsymbol{u}] = 0$ since $\boldsymbol{W}$ is skew symmetric. Hence $\mathrm{RM}(\Omega)$ is in the kernel of the linear elastic potential energy P which implies

$$\alpha \, \|\boldsymbol{u}\|_{H^1(\Omega)} \not\leq J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{u}] = S^{\varepsilon}[\boldsymbol{u}] \leq 2\,|\Omega|, \quad \forall \boldsymbol{u} \in \mathrm{RM}(\Omega) \subset H^1(\Omega).$$

Thus for any unbounded sequence $\{\boldsymbol{u}_k\}_{k \geq 1}$ in $\mathrm{RM}(\Omega)$, the similarity measure S^{ε} remains bounded by $2\,|\Omega|$ while $\|\boldsymbol{u}_k\|_{H^1(\Omega)} \overset{k \to \infty}{\longrightarrow} \infty$. Therefore $J_{\mathrm{reg}}^{\varepsilon}$ is not coercive on $H^1(\Omega)$. However, it can be shown that $J_{\mathrm{reg}}^{\varepsilon}$ is coercive on the orthogonal complement $\mathrm{RM}(\Omega)^{\perp}$ of $\mathrm{RM}(\Omega)$. Lemma 1 in [FKPP10] demonstrates that

$$\mathfrak{H}(\Omega) := \left\{ \boldsymbol{w} \in H^1(\Omega) \,\middle|\, \int_{\Omega} \boldsymbol{w}(\boldsymbol{x}) \, d\boldsymbol{x} = 0, \quad \int_{\Omega} \boldsymbol{w}(\boldsymbol{x})\boldsymbol{x}^{\top} - \boldsymbol{x}\boldsymbol{w}(\boldsymbol{x})^{\top} \, d\boldsymbol{x} = 0 \right\},$$

is an explicit representation of $\mathrm{RM}(\Omega)^{\perp}$ and

$$(8.3) \qquad H^1(\Omega) = \mathrm{RM}(\Omega) \oplus \mathfrak{H}(\Omega).$$

This direct sum decomposition of $H^1(\Omega)$ is crucial for showing well posedness of (8.1). The heuristic idea is to split up the argumentation by considering the minimizations of $J_{\mathrm{reg}}^{\varepsilon}$ on $\mathrm{RM}(\Omega)$ and $\mathfrak{H}(\Omega)$ separately. Then, with the aid of (8.3), the combination of these results implies the existence of a solution to (8.1).

The proofs of both results rely on two observations. First, by defining

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle_P := \frac{1}{2} \int_{\Omega} \lambda (\nabla \cdot \boldsymbol{v}_1)(\nabla \cdot \boldsymbol{v}_2) + \mu (\nabla \boldsymbol{v}_1 + \nabla \boldsymbol{v}_1^{\top}) : (\nabla \boldsymbol{v}_2 + \nabla \boldsymbol{v}_2^{\top}) \, d\boldsymbol{x},$$

for $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathfrak{H}(\Omega)$ and setting

$$|\boldsymbol{v}|_P := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle_P}, \quad \forall \boldsymbol{v} \in \mathfrak{H}(\Omega),$$

one obtains that the linear elastic potential P defines an energy norm on $\mathfrak{H}(\Omega)$. It can be shown that this energy norm is equivalent to the usual $H^1(\Omega)$-norm on $\mathfrak{H}(\Omega)$ [FKPP10, Lem. 2].

Second, following the discussion below Assumption 9, χ_0^{ε} is Lipschitz continuous on $\mathbb{R}^2$. Thus there exists $L > 0$ such that

$$|\chi_0^{\varepsilon}(\boldsymbol{x}) - \chi_0^{\varepsilon}(\boldsymbol{y})| \leq L \, |\boldsymbol{x} - \boldsymbol{y}|, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2.$$

Then for all $\boldsymbol{w}_1, \boldsymbol{w}_2 \in L^2(\Omega)$ we have

$$|S^{\varepsilon}[\boldsymbol{w}_1] - S^{\varepsilon}[\boldsymbol{w}_2]| \leq \int_{\Omega} \left| |\chi_0^{\varepsilon}(\boldsymbol{x} + \boldsymbol{w}_1(\boldsymbol{x})) - \chi_1^{\varepsilon}(\boldsymbol{x})|^2 - |\chi_0^{\varepsilon}(\boldsymbol{x} + \boldsymbol{w}_2(\boldsymbol{x})) - \chi_1^{\varepsilon}(\boldsymbol{x})|^2 \right| \, d\boldsymbol{x}$$

$$\leq 2L \int_{\Omega} |\boldsymbol{w}_1 - \boldsymbol{w}_2| \, d\boldsymbol{x} \leq 2L \sqrt{|\Omega|} \, \|\boldsymbol{w}_1 - \boldsymbol{w}_2\|_{L^2(\Omega)},$$

and thus S^{ε} is continuous on $L^2(\Omega)$. Using equivalence of the norms $|\cdot|_P$ and $\|\cdot\|_{H^1(\Omega)}$ on $\mathfrak{H}(\Omega)$ and continuity of S^{ε} it can be shown that there exists $\boldsymbol{v}^* \in \mathfrak{H}(\Omega)$ such that

$$(8.4) \qquad J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{u} + \boldsymbol{v}^*] = \min_{\boldsymbol{v} \in \mathfrak{H}(\Omega)} J_{\mathrm{reg}}^{\varepsilon}[\boldsymbol{u} + \boldsymbol{v}], \quad \forall \boldsymbol{u} \in \mathrm{RM}(\Omega),$$

compare [FKPP10, Lem. 5]. Thus the minimization of $J_{\mathrm{reg}}^{\varepsilon}$ on $\mathfrak{H}(\Omega)$ is well posed. Showing a similar result on $\mathrm{RM}(\Omega)$ proves to be technically very demanding and significantly more complex. Since $\mathrm{RM}(\Omega)$ is in the kernel of the penalizer P the claim involves solely the similarity measure S^{ε}. The basic idea is to show boundedness of some minimizing sequence of S^{ε} which then implies the existence of a subsequence converging to the minimizer. The most difficult part of the proof is to establish a lower bound for S^{ε} at the limit of the subsequence. As soon as this is achieved, continuity of S^{ε} implies that the limit is indeed the wanted minimizer. In detail, the result states that if for a given $\boldsymbol{w} \in L^2(\Omega)$ there exists $\tilde{\boldsymbol{c}} \in \mathbb{R}^2$ and $\tilde{\boldsymbol{W}} \in \mathbb{S}^2$ such that

$$\int_{\Omega} \left| \chi_0^{\varepsilon}(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) + \tilde{\boldsymbol{c}} + \tilde{\boldsymbol{W}}\boldsymbol{x}) - \chi_1^{\varepsilon}(\boldsymbol{x}) \right|^2 d\boldsymbol{x} < \int_{\Omega} |\chi_1^{\varepsilon}(\boldsymbol{w})|^2 \, d\boldsymbol{x},$$

then there exists $\boldsymbol{c}^* \in \mathbb{R}^2$ and $\boldsymbol{W}^* \in \mathbb{S}^2$ such that

$$(8.5) \qquad S^{\varepsilon}[\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c}^* + \boldsymbol{W}^*\boldsymbol{x}] = \min_{\boldsymbol{c} \in \mathbb{R}^2, \boldsymbol{W} \in \mathbb{S}^2} S^{\varepsilon}[\boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x}],$$

compare [FKPP10, Lem. 4]. Finally, existence of solutions to (8.1) is shown by combining the two auxiliary results covering the existence of minimizing rigid motions and the minimization of $J_{\text{reg}}^{\varepsilon}$ on $\mathfrak{H}(\Omega)$. By (8.3), any $\boldsymbol{w} \in H^1(\Omega)$ can be written in terms of $\boldsymbol{u} \in \text{RM}(\Omega)$ and $\boldsymbol{v} \in \mathfrak{H}(\Omega)$. Thus a minimizing sequence $\{\tilde{\boldsymbol{w}}_k\}_{k \geq 1} \subset H^1(\Omega)$ of $J_{\text{reg}}^{\varepsilon}$ is decomposed into $\{\boldsymbol{u}_k\}_{k \geq 1} \subset \text{RM}(\Omega)$ and $\{\tilde{\boldsymbol{v}}_k\}_{k \geq 1} \subset \mathfrak{H}(\Omega)$. By (8.4), we set $\boldsymbol{v}_k := \arg \min_{\boldsymbol{v} \in \mathfrak{H}(\Omega)} J_{\text{reg}}^{\varepsilon}[\boldsymbol{u}_k + \boldsymbol{v}]$. Thus $\{\boldsymbol{w}_k\}_{k \geq 1}$ defined by $\boldsymbol{w}_k := \boldsymbol{u}_k + \boldsymbol{v}_k$ is also a minimizing sequence of $J_{\text{reg}}^{\varepsilon}$. Then (8.4) and (8.5) is used to argue the existence of accumulation points $\boldsymbol{u}^* \in \text{RM}(\Omega)$ and $\boldsymbol{v}^* \in \mathfrak{H}(\Omega)$ of $\{\boldsymbol{u}_k\}_{k \geq 1}$ and $\{\boldsymbol{v}_k\}_{k \geq 1}$ respectively. Finally $\boldsymbol{w}^* := \boldsymbol{u}^* + \boldsymbol{v}^*$ can be shown to minimize $J_{\text{reg}}^{\varepsilon}$.

Theorem 13. *Given Assumption 9 the optimization problem* (8.1) *has a solution, i.e., there exists* $\boldsymbol{w}^* \in H^1(\Omega)$ *such that*

$$J_{reg}^{\varepsilon}[\boldsymbol{w}^*] = \min_{\boldsymbol{w} \in H^1(\Omega)} J_{reg}^{\varepsilon}[\boldsymbol{w}].$$

Proof. See the proof of Theorem 3 in [FKPP10]. □

Thus the minimization problem (8.1) is indeed well posed for every fixed $\varepsilon > 0$. However, the developed strategy is based on successively decreasing ε to compute a registration. Hence in the next section we address the behavior of solutions to (8.1) as $\varepsilon \to 0$.

8.2 Asymptotic Behavior of Solutions

First, observe that the blurred edge maps χ_i^{ε} converge pointwise to χ_i on Ω as $\varepsilon \to 0$ for $i = 0, 1$ [FKPP10, Lem. 6]. It can be shown that solutions to (8.1) stay "bounded" in a certain sense as $\varepsilon \to 0$. As expected (compare the one-dimensional example discussed in Section 7.3) projections of minimizers onto $\mathfrak{H}(\Omega)$ converge to zero with ε. The infinitesimal rigid motion parts of minimizers, however, do not necessarily tend to zero, although the value of the similarity measure does. Nonetheless, it can be proved that there exists some fixed $\boldsymbol{u}_0 \in \text{RM}(\Omega)$ that has the same limiting property in S^{ε} as the projections onto $\text{RM}(\Omega)$. In other words, the projections onto $\text{RM}(\Omega)$ do not grow infinitely as $\varepsilon \to 0$.

Theorem 14. *Given Assumption 9 let* $\boldsymbol{w}_{\varepsilon} \in H^1(\Omega)$ *be a solution to* (8.1) *for* $\varepsilon > 0$ *and denote by* $\boldsymbol{u}_{\varepsilon} \in RM(\Omega)$ *and* $\boldsymbol{v}_{\varepsilon} \in \mathfrak{H}(\Omega)$ *the projections of* $\boldsymbol{w}_{\varepsilon}$ *onto* $RM(\Omega)$ *and* $\mathfrak{H}(\Omega)$ *respectively such that* $\boldsymbol{w}_{\varepsilon} = \boldsymbol{u}_{\varepsilon} + \boldsymbol{v}_{\varepsilon}$. *Then* $\boldsymbol{v}_{\varepsilon} \xrightarrow[H^1(\Omega)]{\varepsilon \to 0} 0$ *and there exists* $\boldsymbol{u}_0 \in RM(\Omega)$ *satisfying*

$$\lim_{\varepsilon \to 0} S^{\varepsilon}[\boldsymbol{u}_0 + \boldsymbol{v}_{\varepsilon}] = \lim_{\varepsilon \to 0} S^{\varepsilon}[\boldsymbol{u}_{\varepsilon} + \boldsymbol{v}_{\varepsilon}] = 0.$$

Proof. See the proof of Theorem 4 in [FKPP10]. □

Recall that the motivation for introducing the blurred edge maps χ_i^{ε} was to enable the use of the SSID measure (7.1) despite the fact that the underlying edge sets have Lebesgue measure zero. As mentioned above, the Hausdorff distance (7.2) on the other hand is explicitly designed to measure differences in lower dimensional objects such as edges. This raises the question quite naturally whether there is a relation connecting the Hausdorff distance between Γ_0 and Γ_1 and the augmented SSID measure S^{ε} of the respective blurred edge maps χ_0^{ε} and χ_1^{ε}. It turns out that the framework of shape analysis can be used to
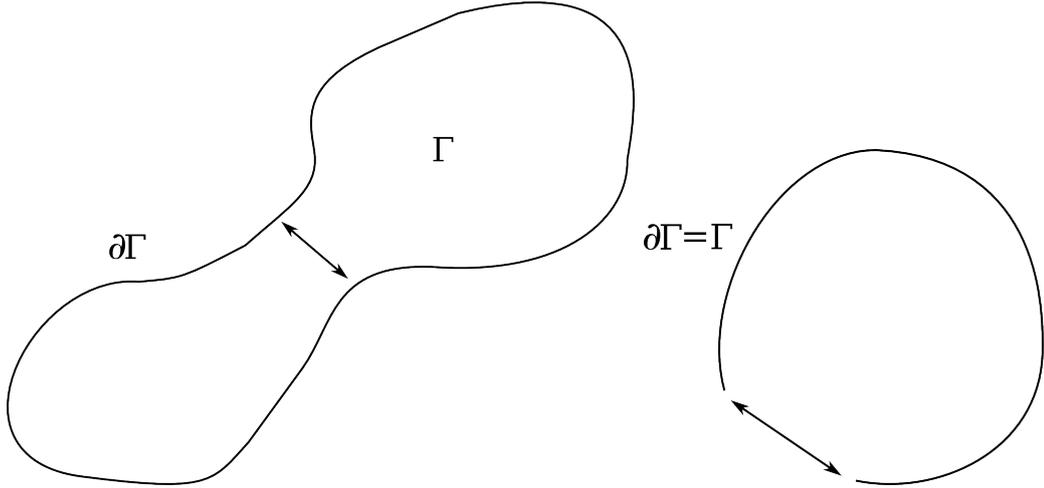
Figure 8.1: Two examples of admissible shapes. Any regular curve with curvature bounded from above by $\frac{1}{h_0}$ and pinch distance (symbolized by arrows in the sketches) bounded from below by h_0 is in $\mathcal{S}$. Shown is a closed (left) and an open curve (right). Sketches are patterned on Figure 4 in [CFK04].

establish a convergence criterion. For the following we rely on techniques presented in [CFK04]. We start by considering a set of smooth shapes. We introduce

$$\mathcal{C}_2 := \{\Gamma \subset \Omega \,|\, \partial\Gamma \neq \emptyset \text{ and } \partial\Gamma \text{ can be locally represented}$$
$$\text{as an epigraph of a twice differentiable function}\} \,.$$

In other words the boundary of $\Gamma \in \mathcal{C}_2$ is a "simple" regular curve. Nonetheless, shapes in $\mathcal{C}_2$ may still bend and pinch arbitrarily. However, in order to establish limiting properties of shape sequences, additional regularity is required. Thus we further introduce the *h-tubular neighborhood of a shape* Γ

$$U_h(\Gamma) := \{\boldsymbol{y} \in \Omega \,|\, d_\Gamma(\boldsymbol{y}) < h\} \,,$$

for $h > 0$. Additionally, let

$$\Pi_\Gamma(\boldsymbol{x}) := \{\boldsymbol{p} \in \bar{\Gamma} \,|\, |\boldsymbol{p} - \boldsymbol{x}| = d_\Gamma(\boldsymbol{x})\} \,,$$

denote the set of projections of $\boldsymbol{x} \in \Omega$ on Γ. Assume there exists $h > 0$ such that $\Pi_\Gamma(\boldsymbol{x})$ is a singleton for every $\boldsymbol{x} \in U_h(\Gamma)$. The maximal value for which this property holds is called the *reach of* Γ, noted reach(Γ). For given fixed $h_0 > 0$ we define $\mathcal{F}_{h_0}$ as the collection of all sets with reach(Γ) $\geq h_0$, i.e.,

$$\mathcal{F}_{h_0} := \{\Gamma \subset \Omega \,|\, \text{reach}(\Gamma) \geq h_0\} \,.$$

Now we can define the set $\mathcal{S}$ of all admissible shapes as

$$\mathcal{S} := \mathcal{C}_2 \cap \mathcal{F}_{h_0}.$$

Thus $\mathcal{S}$ consists of shapes which have a boundary that is regular in terms of $\mathcal{C}_2$ and has

constrained reach by means of $\mathcal{F}_{h_0}$. Figure 8.1 shows two exemplary shapes in $\mathcal{S}$. Elements of $\mathcal{S}$ now have sufficient structure to formulate the following convergence criterion. Let $\{\Gamma_k\}_{k\geq 1}$ be a sequence in $\mathcal{S}$ and assume $\Gamma^* \in \mathcal{S}$. Then it can be shown [CFK04, Prop. 4, Prop. 5] that

$$(8.6) \qquad d_{\Gamma_k} \underset{H^1(\Omega)}{\longrightarrow} d_{\Gamma^*} \quad \Leftrightarrow \quad d_{\mathrm{H}}(\Gamma_k, \Gamma^*) \overset{k\to\infty}{\longrightarrow} 0.$$

To embed this result in the registration framework discussed here suppose $\Gamma_k = \mathrm{supp}(\chi_0 \circ (\mathrm{id} + \boldsymbol{w}_k))$ and $\Gamma^* = \mathrm{supp}(\chi_1)$, where χ_0 denotes the template and χ_1 the reference edge map. Note carefully that $\boldsymbol{w}_k$ has to be sufficiently smooth to have $\Gamma_k \in \mathcal{S}$. For fixed $\varepsilon > 0$ define the respective mollified edge maps χ_i^ε according to (7.3) for $i = 0, 1$. Further, using the introduced notation let $\bar{U}_\varepsilon(\Gamma_k)$ and $\bar{U}_\varepsilon(\Gamma^*)$ be the closed ε-tubular neighborhoods of Γ_k and Γ^* respectively. Note that by setting $\boldsymbol{w}_0 := 0$ this yields by Definition (7.3) that $\mathrm{supp}(\chi_0^\varepsilon) = \bar{U}_\varepsilon(\Gamma_0)$ and similarly $\mathrm{supp}(\chi_1^\varepsilon) = \bar{U}_\varepsilon(\Gamma^*)$. Suppose $d_{\mathrm{H}}(\Gamma_k, \Gamma^*) \overset{k\to\infty}{\longrightarrow} 0$ then (8.6) implies

$$\begin{aligned}
0 \overset{\infty\leftarrow k}{\longleftarrow} \frac{1}{2\varepsilon} \|d_{\Gamma_k} - d_{\Gamma^*}\|_{H^1(\Omega)}^2 &\geq \frac{1}{2\varepsilon} \|d_{\Gamma_k} - d_{\Gamma^*}\|_{L^2(\Omega)}^2 \\
&\geq \frac{1}{2} \int_{\bar{U}_\varepsilon(\Gamma_k) \cup \bar{U}_\varepsilon(\Gamma^*)} \left| \frac{d_{\Gamma_k}}{\varepsilon} - \frac{d_{\Gamma^*}}{\varepsilon} \right|^2 d\boldsymbol{x} \\
&= \frac{1}{2} \int_\Omega |\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}_k) - \chi_1^\varepsilon|^2 d\boldsymbol{x} = S^\varepsilon[\boldsymbol{w}_k].
\end{aligned}$$

In other words, if the reference and the template edge sets are sufficiently regular, i.e., $\Gamma_0, \Gamma^* \in \mathcal{S}$, and the reference edge set Γ^* can be expressed as the Hausdorff limit of a sequence of smoothly deformed edge sets Γ_k, then the introduced similarity measure S^ε converges to zero. Thus the Hausdorff distance is a stronger distance measure than S^ε. Note carefully that this weaker notion of similarity in S^ε is expressly desired in our context because of the following. Due to the sensitivity of the Hausdorff distance small deviations in the template and reference edge sets are clearly reflected in $d_{\mathrm{H}}(\Gamma_0, \Gamma^*)$. Thus compared to minimizing a SSID type measure, $H^1(\Omega)$-similarity or by (8.6) minimal Hausdorff distance for admissible shapes, may require substantial algorithmic effort while possibly only slightly changing the computed deformation field. Moreover, in practice we usually cannot guarantee that template and reference edge sets are admissible shapes. Then a minimization of the Hausdorff distance might be even more complex.

In contrast our developed strategy (see Algorithm 9.1) is based on the idea of computing a rough registration for large values of $\varepsilon > 0$ that is iteratively refined as ε becomes smaller. Thus $H^1(\Omega)$-similarity is not wanted for large values of ε and not needed for small ε since we favor local minimizers of $J_{\mathrm{reg}}^\varepsilon$ (compare the discussion of the one-dimensional example in Section 7.3).

Having discussed existence and asymptotic properties of solutions to (8.1) the next section sets up the associated optimality system and establishes a solution strategy.

8.3 Optimality Conditions and Solution Strategy

To derive an optimality system for the minimization of $J_{\mathrm{reg}}^\varepsilon$ we use again Gâteaux differential calculus (compare Appendix B.2). Thus we start by noting that $J_{\mathrm{reg}}^\varepsilon$ is everywhere

Gâteaux differentiable. For the sake of a clear presentation we compute the derivatives of S^ε and P separately. The Gâteaux derivative of the similarity measure is readily established

$$\frac{\delta S^\varepsilon}{\delta \boldsymbol{w}}[\boldsymbol{w}; \boldsymbol{v}] = \frac{d}{ds}\left(\frac{1}{2}\int_\Omega |\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w} + s\boldsymbol{v}) - \chi_1^\varepsilon|^2\, d\boldsymbol{x}\right)\bigg|_{s=0}$$

$$= \int_\Omega (\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w} + s\boldsymbol{v}) - \chi_1^\varepsilon)\, \nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w} + s\boldsymbol{v}) \cdot \boldsymbol{v}\, d\boldsymbol{x}\bigg|_{s=0}$$

$$= \int_\Omega (\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) - \chi_1^\varepsilon)\, \nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \cdot \boldsymbol{v}\, d\boldsymbol{x}.$$

For the linear elastic potential we obtain

$$\frac{\delta P}{\delta \boldsymbol{w}}[\boldsymbol{w}; \boldsymbol{v}] = \frac{d}{ds}\left(\frac{1}{2}\int_\Omega \lambda(\nabla \cdot (\boldsymbol{w} + s\boldsymbol{v}))^2 + \mu\left|\nabla(\boldsymbol{w} + s\boldsymbol{v})^\top + \nabla(\boldsymbol{w} + s\boldsymbol{v})\right|^2 d\boldsymbol{x}\right)\bigg|_{s=0}$$

$$= \int_\Omega \lambda(\nabla \cdot \boldsymbol{w})(\nabla \cdot \boldsymbol{v}) + \mu(\nabla \boldsymbol{w}^\top + \nabla \boldsymbol{w}) : (\nabla \boldsymbol{v}^\top + \nabla \boldsymbol{v})\, d\boldsymbol{x},$$

and thus the weak formulation of the necessary optimality condition associated to (8.1) is given by (compare, e.g., [Lue69, Sec. 7.4, Th. 1])

$$(8.7) \qquad 0 = \frac{\delta J_{\mathrm{reg}}^\varepsilon}{\delta \boldsymbol{w}}[\boldsymbol{w}; \boldsymbol{v}] = \frac{\delta S^\varepsilon}{\delta \boldsymbol{w}}[\boldsymbol{w}; \boldsymbol{v}] + \frac{\delta P}{\delta \boldsymbol{w}}[\boldsymbol{w}; \boldsymbol{v}], \quad \forall \boldsymbol{v} \in H^1(\Omega).$$

Note that the composition of χ_0^ε with $\boldsymbol{w}$ in $\frac{\delta S^\varepsilon}{\delta \boldsymbol{w}}$ makes the derived weak formulation nonlinear in $\boldsymbol{w}$. Thus in contrast to the weak problems discussed so far, (8.7) cannot be written in terms of a bilinear form and a linear functional and hence requires a different solution strategy (for an introduction to non-linear problems see, e.g., [Eva08, Pt. III]). One approach to solve such a non-linear problem is to employ an infinite dimensional version of Newton's method (for an exhaustive review see [Deu04, Chap. 8]). Thus we introduce

$$(8.8) \qquad \begin{cases} \dfrac{\delta^2 J_{\mathrm{reg}}^\varepsilon}{\delta \boldsymbol{w}^2}[\boldsymbol{w}_k; \boldsymbol{v}, \boldsymbol{d}_k] = -\dfrac{\delta J_{\mathrm{reg}}^\varepsilon}{\delta \boldsymbol{w}}[\boldsymbol{w}_k; \boldsymbol{v}], \quad \forall \boldsymbol{v} \in C^\infty(\bar{\Omega}), \\[2mm] \boldsymbol{w}_{k+1} = \boldsymbol{w}_k + t\boldsymbol{d}_k, \end{cases} \quad k = 1, 2, \ldots,$$

where $t > 0$ denotes a given step size. Note carefully that (8.8) is linear in $\boldsymbol{d}_k$. Thus by considering (8.8) instead of (8.7) we transform the solution of one non-linear equation to the iterative solution of several linear problems. For computing the second Gâteuax derivative of $J_{\mathrm{reg}}^\varepsilon$ we split up the calculation again and start by considering

$$\frac{\delta^2 P}{\delta \boldsymbol{w}^2}[\boldsymbol{w}; \boldsymbol{v}, \boldsymbol{d}] = \frac{d}{ds}\left(\int_\Omega \lambda(\nabla \cdot (\boldsymbol{w} + s\boldsymbol{d}))(\nabla \cdot \boldsymbol{v})\, d\boldsymbol{x}\right.$$

$$\left.+ \int_\Omega \mu(\nabla(\boldsymbol{w} + s\boldsymbol{d})^\top + \nabla(\boldsymbol{w} + s\boldsymbol{d})) : (\nabla \boldsymbol{v}^\top + \nabla \boldsymbol{v})\, d\boldsymbol{x}\right)\bigg|_{s=0}$$

$$= \int_\Omega \lambda(\nabla \cdot \boldsymbol{d})(\nabla \cdot \boldsymbol{v}) + \mu(\nabla \boldsymbol{d}^\top + \nabla \boldsymbol{d}) : (\nabla \boldsymbol{v}^\top + \nabla \boldsymbol{v})\, d\boldsymbol{x}.$$

For the second derivative of the similarity measure we get

$$
\frac{\delta^2 S^\varepsilon}{\delta \boldsymbol{w}^2}[\boldsymbol{w}; \boldsymbol{v}, \boldsymbol{d}] = \frac{d}{ds} \left(\int_\Omega (\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w} + s\boldsymbol{d}) - \chi_1^\varepsilon) \, \nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w} + s\boldsymbol{d}) \cdot \boldsymbol{v} \, d\boldsymbol{x} \right) \bigg|_{s=0}
$$

(8.9)
$$
= \int_\Omega \boldsymbol{v}^\top \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \right) \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \right)^\top \boldsymbol{d} \, d\boldsymbol{x}
$$

$$
+ \int_\Omega (\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) - \chi_1^\varepsilon) \, \boldsymbol{v}^\top \left(\nabla^2 \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \right) \boldsymbol{d} \, d\boldsymbol{x}.
$$

Note that the second non-symmetric term in (8.9) disturbs the positive definiteness of $\frac{\delta^2 J_{\mathrm{reg}}^\varepsilon}{\delta \boldsymbol{w}^2}$. Neglecting it we obtain an approximation of $\frac{\delta^2 S^\varepsilon}{\delta \boldsymbol{w}^2}$ that only involves the symmetrized gradient of the blurred template edge map

$$
\frac{\delta^2 S^\varepsilon}{\delta \boldsymbol{w}^2}[\boldsymbol{w}; \boldsymbol{v}, \boldsymbol{d}] \approx \int_\Omega \boldsymbol{v}^\top \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \right) \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \right)^\top \boldsymbol{d} \, d\boldsymbol{x}.
$$

Thus we can write an approximation of the second derivative of $J_{\mathrm{reg}}^\varepsilon$ in terms of a bilinear form $B_{\mathrm{reg}} : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ given by

$$
B_{\mathrm{reg}}[\boldsymbol{d}, \boldsymbol{v}] := \int_\Omega \lambda (\nabla \cdot \boldsymbol{d})(\nabla \cdot \boldsymbol{v}) + \mu (\nabla \boldsymbol{d}^\top + \nabla \boldsymbol{d}) : (\nabla \boldsymbol{v}^\top + \nabla \boldsymbol{v}) \, d\boldsymbol{x}
$$

$$
+ \int_\Omega \boldsymbol{v}^\top \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \right) \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \right)^\top \boldsymbol{d} \, d\boldsymbol{x},
$$

for some $\boldsymbol{w} \in H^1(\Omega)$. Note that for a given deformation field $\boldsymbol{w}$ the right hand side of (8.8) is a linear function of $\boldsymbol{v} \in H^1(\Omega)$. Thus we introduce $f_{\mathrm{reg}} : H^1(\Omega) \to \mathbb{R}$ defined by

$$
f_{\mathrm{reg}}[\boldsymbol{v}] := - \int_\Omega \lambda (\nabla \cdot \boldsymbol{w})(\nabla \cdot \boldsymbol{v}) + \mu (\nabla \boldsymbol{w}^\top + \nabla \boldsymbol{w}) : (\nabla \boldsymbol{v}^\top + \nabla \boldsymbol{v}) \, d\boldsymbol{x}
$$

$$
- \int_\Omega (\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) - \chi_1^\varepsilon) \, \nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}) \cdot \boldsymbol{v} \, d\boldsymbol{x}.
$$

Thus the Newton step seen in (8.8) is approximated by the following variational problem

(8.10)
$$
B_{\mathrm{reg}}[\boldsymbol{d}, \boldsymbol{v}] = f_{\mathrm{reg}}[\boldsymbol{v}], \quad \forall \boldsymbol{v} \in H^1(\Omega).
$$

Note that straight-forward calculations show boundedness of B_{reg} and f_{reg}. If the deformed blurred template edge map $\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w})$ is "sufficiently unsymmetric" then by using Korn's inequality [Bre04] it can be proved that the second term in B_{reg} coerces the kernel of the first term. Then the Lax–Milgram Lemma (Appendix (B.2)) yields existence and uniqueness of a solution to (8.10). Thus each approximate Newton step is well posed. We summarize this fact in the following result.

Theorem 15. *Given Assumption 9 let $\boldsymbol{w} \in H^1(\Omega)$. Suppose*

$$
\int_\Omega |\nabla \chi_0^\varepsilon \circ (id + \boldsymbol{w}) \cdot (\boldsymbol{c} + \boldsymbol{W}\boldsymbol{x})|^2 \, d\boldsymbol{x} = 0 \ \text{implies} \ \boldsymbol{c} + \boldsymbol{W}\boldsymbol{x} = 0, \quad \forall \boldsymbol{W} \in \mathbb{S}^2, \forall \boldsymbol{c} \in \mathbb{R}^2.
$$

Then there exists a unique $\boldsymbol{d} \in H^1(\Omega)$ satisfying (8.10).

Proof. See the proof of Theorem 2 in [FKPP10]. □

Under standard assumptions convergence of Newton's method can be proved using classical results such as the Newton–Kantorovich Theorem (see, e.g., [Deu04]). For solving (8.10) in practice we consider the following. Note that while (8.10) is linear in $\boldsymbol{d}$, both B_{reg} and f_{reg} depend on the deformed blurred template edge map $\chi_0^\varepsilon \circ (\text{id} + \boldsymbol{w})$ or explicitly $\chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}))$. Hence $\boldsymbol{w}$ appears as argument of χ_0^ε which complicates a finite element discretization of (8.10) considerably as opposed to a standard finite difference approach. Note further that in the present context it is uncertain whether a finite element approach poses any significant improvements over a finite difference formulation. In contrast, the practical realization of the Inner and Outer Iterations (discussed in Chapter 5) greatly benefits from the use of finite elements; since it is not clear how to discretize consistently the binary edge map χ using finite differences. However, in the context of finite elements, χ is simply a variable coefficient in a bilinear form. Thus we employed a finite element approach to establish an unambiguous discretization of the binary edge map χ. Conversely, the registration framework discussed here is based on blurred edge maps which can be discretized naturally using finite differences. Furthermore, finite difference approximations are thoroughly investigated and well established in elastic image registration (compare [Mod04, Sec. 9.4]). Hence we rely on a finite difference discretization of (8.10). Thus we start by deducing the associated strong formulation of (8.7) by first shifting derivatives from $\boldsymbol{v}$ to $\boldsymbol{d}$ in $\frac{\delta P}{\delta \boldsymbol{w}}$ using partial integration

$$(8.11) \qquad \frac{\delta P}{\delta \boldsymbol{w}}[\boldsymbol{w}; \boldsymbol{v}] = \int_{\partial\Omega} \lambda(\nabla \cdot \boldsymbol{w})(\boldsymbol{v} \cdot \boldsymbol{n}) + \mu \boldsymbol{v}^\top (\nabla \boldsymbol{w}^\top + \nabla \boldsymbol{w})\boldsymbol{n} \, dS$$
$$- \int_{\Omega} \lambda \nabla(\nabla \cdot \boldsymbol{w}) \cdot \boldsymbol{v} + \mu(\nabla(\nabla \cdot \boldsymbol{w}) + \Delta \boldsymbol{w}) \cdot \boldsymbol{v} \, d\boldsymbol{x}.$$

We use this representation of $\frac{\delta P}{\delta \boldsymbol{w}}$ in (8.7). By density of $C^\infty(\bar{\Omega})$ in $H^1(\Omega)$ [Ada75, Th. 3.16], (8.7) holds for any variation $\boldsymbol{v} \in C^\infty(\bar{\Omega})$. Thus we may apply the fundamental Lemma of calculus of variations [Ada75, Cor. 3.26] to obtain the strong formulation

$$(8.12) \qquad \begin{cases} \mathcal{E}\boldsymbol{w} = \boldsymbol{f}[\boldsymbol{w}], & \text{in } \Omega, \\ \lambda n_l \nabla \cdot \boldsymbol{w} + \mu(\nabla w_l + \dfrac{\partial}{\partial x_l}\boldsymbol{w}) \cdot \boldsymbol{n} = 0, & \text{on } \partial\Omega, \quad l = 1, 2, \end{cases}$$

with

$$(8.13) \qquad \mathcal{E}\boldsymbol{w} := (\lambda + \mu)\nabla(\nabla \cdot \boldsymbol{w}) + \mu\Delta\boldsymbol{w},$$

and

$$(8.14) \qquad \boldsymbol{f}[\boldsymbol{w}] := (\chi_0^\varepsilon \circ (\text{id} + \boldsymbol{w}) - \chi_1^\varepsilon) \nabla \chi_0^\varepsilon \circ (\text{id} + \boldsymbol{w}),$$

where $\boldsymbol{w}(\boldsymbol{x}) := (w_1(\boldsymbol{x}), w_2(\boldsymbol{x}))$ and $\boldsymbol{n} := (n_1, n_2)$ denotes the outer unit normal vector on $\partial\Omega$. Note that the boundary condition in (8.12) results from letting $\boldsymbol{v}$ be a smooth function that is concentrated on $\partial\Omega$ such that the second integral in (8.11) vanishes more rapidly than the boundary integral and otherwise $\boldsymbol{v}$ has only one locally non-trivial component. The operator $\mathcal{E}$ is the so-called *elasticity operator* and $\boldsymbol{f}$ is the *driving force* of the registration. Note that (8.12) are the Euler–Lagrange equations associated to the minimization problem (8.1), i.e., the strong formulation of (8.7). Thus (8.12) is similarly a system of

non-linear PDEs. Observe further that the driving force $\boldsymbol{f}$ is the strong formulation of the derivative of the similarity measure S^ε. The same correspondence exists between the elasticity operator $\mathcal{E}$ and the linear elastic penalty P.

Analogous to the calculation carried out in (8.11) we also apply partial integration in $\frac{\delta^2 P}{\delta \boldsymbol{w}^2}$. Then we may rely again on the fundamental Lemma of calculus of variations to derive a strong formulation of the Newton step (8.10). By (8.8), $\boldsymbol{d}_k = \frac{1}{t}(\boldsymbol{w}_{k+1} - \boldsymbol{w}_k)$, thus if $\boldsymbol{w}_k$ satisfies the boundary conditions of (8.12) so does $\boldsymbol{d}_k$ and hence we obtain

$$(8.15) \quad \begin{cases} \left(-\mathcal{E} + \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}_k) \right) \left(\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w}_k) \right)^\top \right) \boldsymbol{d}_k = \mathcal{E} \boldsymbol{w}_k - \boldsymbol{f}[\boldsymbol{w}_k], & \text{in } \Omega, \\ \lambda n_l \nabla \cdot \boldsymbol{w}_k + \mu(\nabla w_{k_l} + \dfrac{\partial}{\partial x_l} \boldsymbol{w}_k) \cdot \boldsymbol{n} = 0, & \text{on } \partial\Omega, \end{cases}$$

for $l = 1, 2$ and $\boldsymbol{w}_k(\boldsymbol{w}) := (w_{k_1}(\boldsymbol{x}), w_{k_2}(\boldsymbol{x}))$ with $k \in \mathbb{N}$. Note that (8.15) as being the strong formulation of (8.8) is a linear PDE system in $\boldsymbol{d}_k$.

In the next chapter we set up a finite difference discretization of (8.15) and give a detailed pseudo-code of the proposed registration approach for edge maps.

9

Practical Realization

In this chapter we develop an explicit representation of the the strong formulation of the Newton step (8.15) in the discrete setting using finite differences. Following the presentation in [FKPP12] we subsequently set up a detailed pseudo-code of the proposed strategy for the elastic registration of edge maps. Finally, we embed the novel approach for computing binary edge maps introduced in Chapter 3 in the registration framework discussed here.

9.1 Discretization of the Newton Step

In agreement with terminology introduced in Chapter 5 let $h = 1/N$ denote the cellsize used and accordingly let Ω_h be an equidistant grid of size $N \times N$ on Ω with cellwidth h. For the following derivation we drop the iteration index k seen in (8.15) to avoid excessively complex notation. We start by discretizing the elasticity operator $\mathcal{E}$ given by (8.13). Thus let $\boldsymbol{x} = (x, y) \in \Omega$ and $\boldsymbol{w} := (u, v)$. We rewrite (8.13) more explicitly in terms of the component functions u and v of $\boldsymbol{w}$, i.e.,

$$\mathcal{E}\boldsymbol{w} = (\lambda + \mu)\nabla\left(\frac{\partial}{\partial x}u + \frac{\partial}{\partial y}v\right) + \mu\left(\frac{\partial^2}{\partial x^2}\boldsymbol{w} + \frac{\partial^2}{\partial y^2}\boldsymbol{w}\right)$$
$$= \begin{pmatrix} (\lambda + 2\mu)\frac{\partial^2}{\partial x^2} + \mu\frac{\partial^2}{\partial y^2} & (\lambda + \mu)\frac{\partial^2}{\partial x \partial y} \\ (\lambda + \mu)\frac{\partial^2}{\partial x \partial y} & \mu\frac{\partial^2}{\partial x^2} + (\lambda + 2\mu)\frac{\partial^2}{\partial y^2} \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}.$$

With operators $\mathcal{E}^{i,j}$, $1 \le i, j \le 2$, acting on u and v this can be rewritten as

$$\mathcal{E}\boldsymbol{w} = \begin{pmatrix} \mathcal{E}^{1,1} & \mathcal{E}^{1,2} \\ \mathcal{E}^{2,1} & \mathcal{E}^{2,2} \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}.$$

Let $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{N^2}$ be lexicographically ordered approximations of u and v respectively on the grid Ω_h. Then we can associate matrices $\boldsymbol{E}^{i,j} \in \mathbb{R}^{N^2 \times N^2}$ to the operators $\mathcal{E}^{i,j}$ such that

$$\mathcal{E}[u(\Omega_h), v(\Omega_h)] \approx \begin{pmatrix} \boldsymbol{E}^{1,1} & \boldsymbol{E}^{1,2} \\ \boldsymbol{E}^{2,1} & \boldsymbol{E}^{2,2} \end{pmatrix}\begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix}.$$

The stencils for the matrices are given in Appendix C.3. Then with

$$\boldsymbol{E} := \begin{pmatrix} \boldsymbol{E}^{1,1} & \boldsymbol{E}^{1,2} \\ \boldsymbol{E}^{2,1} & \boldsymbol{E}^{2,2} \end{pmatrix} \in \mathbb{R}^{2N^2 \times 2N^2} \quad \text{and} \quad \boldsymbol{W} := \begin{pmatrix} \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \in \mathbb{R}^{2N^2},$$

we can write this block matrix system as

$$\boldsymbol{EW} = \left(\begin{array}{cc} \boldsymbol{E}^{1,1} & \boldsymbol{E}^{1,2} \\ \boldsymbol{E}^{2,1} & \boldsymbol{E}^{2,2} \end{array} \right) \left(\begin{array}{c} \boldsymbol{U} \\ \boldsymbol{V} \end{array} \right).$$

The discretization of terms involving the deformed blurred template edge map $\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w})$ requires the use of an interpolation scheme if $\boldsymbol{w}$ contains non-integer values. We use bilinear interpolation (see, e.g., [Mod04, Sec. 3.1.3]) to approximate $\chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}))$ if $\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})$ is not a grid point and assign the extrapolation value zero if $\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}) \notin \Omega$. Thus let $\boldsymbol{\chi}_0^\varepsilon \in \mathbb{R}^{N^2}$ denote the lexicographically ordered vector that approximates $\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w})$ on Ω_h according to this procedure. The finite difference discretization of $\nabla \chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w})$ is obtained analogously. We consider the partial derivatives of $\chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}))$ separately. Thus we introduce the vectors $\mathfrak{D}_x \boldsymbol{\chi}_0^\varepsilon \in \mathbb{R}^{N^2}$ and $\mathfrak{D}_y \boldsymbol{\chi}_0^\varepsilon \in \mathbb{R}^{N^2}$ holding central difference approximations of $\frac{\partial}{\partial x} \chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}))$ and $\frac{\partial}{\partial y} \chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}))$ respectively on Ω_h. In accordance with Section 5.3, let $\boldsymbol{\mathcal{D}}[\boldsymbol{Y}]$ denote again the diagonal matrix with the values of some vector $\boldsymbol{Y}$ on its main diagonal. Then a finite difference approximation of the symmetrized gradient of $\chi_0^\varepsilon \circ (\mathrm{id} + \boldsymbol{w})$ on the left hand side of (8.15) is given by

$$\mathfrak{D}\boldsymbol{\chi}_0^\varepsilon := \left(\begin{array}{cc} \boldsymbol{\mathcal{D}}[\mathfrak{D}_x \boldsymbol{\chi}_0^\varepsilon]\boldsymbol{\mathcal{D}}[\mathfrak{D}_x \boldsymbol{\chi}_0^\varepsilon] & \boldsymbol{\mathcal{D}}[\mathfrak{D}_x \boldsymbol{\chi}_0^\varepsilon]\boldsymbol{\mathcal{D}}[\mathfrak{D}_y \boldsymbol{\chi}_0^\varepsilon] \\ \boldsymbol{\mathcal{D}}[\mathfrak{D}_y \boldsymbol{\chi}_0^\varepsilon]\boldsymbol{\mathcal{D}}[\mathfrak{D}_x \boldsymbol{\chi}_0^\varepsilon] & \boldsymbol{\mathcal{D}}[\mathfrak{D}_y \boldsymbol{\chi}_0^\varepsilon]\boldsymbol{\mathcal{D}}[\mathfrak{D}_y \boldsymbol{\chi}_0^\varepsilon] \end{array} \right) \in \mathbb{R}^{2N^2 \times 2N^2}.$$

Thus $\mathfrak{D}\boldsymbol{\chi}_0^\varepsilon$ has the same size as $\boldsymbol{E}$ and is comprised of diagonal blocks. Note that the multiplication of two diagonal matrices corresponds to situating a componentwise vector multiplication on the diagonal of a matrix.

A finite difference discretization of the driving force $\boldsymbol{f}$ given by (8.14) can be established in a similar fashion. Considering the partial derivatives of $\chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x}))$ again separately we have

$$\boldsymbol{f}[\boldsymbol{w}] = \left(\begin{array}{c} (\chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})) - \chi_1^\varepsilon(\boldsymbol{x})) \frac{\partial}{\partial x} \chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})) \\ (\chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})) - \chi_1^\varepsilon(\boldsymbol{x})) \frac{\partial}{\partial y} \chi_0^\varepsilon(\boldsymbol{x} + \boldsymbol{w}(\boldsymbol{x})) \end{array} \right).$$

The discretization of the blurred reference edge map is achieved in a straight forward manner. Thus let $\boldsymbol{\chi}_1^\varepsilon \in \mathbb{R}^{N^2}$ denote the vector of values of χ_1^ε on Ω_h in lexicographic ordering. Then using the notation introduced above the driving force $\boldsymbol{f}$ is discretized by

$$\boldsymbol{F} := \left(\begin{array}{c} \boldsymbol{\mathcal{D}}[\boldsymbol{\chi}_0^\varepsilon - \boldsymbol{\chi}_1^\varepsilon]\mathfrak{D}_x \boldsymbol{\chi}_0^\varepsilon \\ \boldsymbol{\mathcal{D}}[\boldsymbol{\chi}_0^\varepsilon - \boldsymbol{\chi}_1^\varepsilon]\mathfrak{D}_y \boldsymbol{\chi}_0^\varepsilon \end{array} \right) \mathbb{R}^{2N^2}.$$

Thus, in terms of the introduced expressions a finite difference discretization of the Newton step is established as follows. Find $\boldsymbol{D} \in \mathbb{R}^{2N^2}$ solving

$$(-\boldsymbol{E} + \mathfrak{D}\boldsymbol{\chi}_0^\varepsilon)\, \boldsymbol{D} = -\boldsymbol{EW} - \boldsymbol{F}.$$

Note that for the sake of readability the introduced symbols $\mathfrak{D}\boldsymbol{\chi}_0^\varepsilon$ and $\boldsymbol{F}$ do not reflect the dependence of the respective quantities on the deformation field $\boldsymbol{w}$. However, in the course of deriving a discrete version of Newton's method we have to identify variables explicitly that change in the course of the iteration. Hence we take up the iteration index k again and augment all symbols that depend upon the current deformation field by the

subscript k. Then a finite difference discretization of (8.15) takes the form

$$(9.1) \quad \begin{cases} \left(-\boldsymbol{E} + \boldsymbol{\mathfrak{D}}\boldsymbol{\chi}_{0,k}^{\varepsilon}\right)\boldsymbol{D}_k = -\boldsymbol{E}\boldsymbol{W}_k - \boldsymbol{F}_k, \\ \boldsymbol{W}_{k+1} = \boldsymbol{W}_k + t_k\boldsymbol{D}_k, \end{cases} \quad k = 1, 2, \ldots,$$

with $t_k > 0$. Having set up the discrete framework for (8.8) we can now focus on the details of the registration strategy.

9.2 A Registration Strategy for Edge Maps

Based on the discrete approximations introduced in the previous section we start our discussion by considering the *line-search strategy* used, i.e., the technique used to determine the step size t_k. Thus let $J_{\mathrm{reg},h}^{\varepsilon}$ denote a discrete approximation of $J_{\mathrm{reg}}^{\varepsilon}$ and let $L > 1$ be a natural number. Then t_k is computed by solving

$$(9.2) \quad \begin{cases} t_k = \arg\min_{t \in \mathfrak{T}} J_{\mathrm{reg},h}^{\varepsilon}[\boldsymbol{W}_k + t\boldsymbol{D}_k], \\ \mathfrak{T} := \left\{ t = \dfrac{2l}{L} \,\middle|\, l = 1, \ldots, L \right\}. \end{cases}$$

In contrast to common line-search algorithms found in literature, e.g., Armijo–Goldstein or Wolfe–Powell strategies (compare [NW00, Chap. 3] or [Kel99, Sec. 3.2]), (9.2) explicitly allows step sizes larger than one by minimizing $J_{\mathrm{reg},h}^{\varepsilon}$ on the interval $[2/L, 2]$. In numerous practical tests the approach (9.2) proved to be very reliable and computationally efficient. As long as L is not chosen exceptionally large, only a few evaluations of $J_{\mathrm{reg},h}^{\varepsilon}$ are needed to solve (9.2).

For establishing appropriate stopping criteria we make the following crucial observation. Note that the right hand side of (9.1) is the discrete formulation of the Euler–Lagrange equations (8.12). Since the Newton iteration was developed to compute a solution of (8.1) we ideally want that $\boldsymbol{E}\boldsymbol{W} = \boldsymbol{F}$. We use this property to formulate a relative stopping criterion. Let $r_{k+1} := -\boldsymbol{E}\boldsymbol{W}_{k+1} - \boldsymbol{F}_{k+1}$ and define

$$(9.3) \quad r_b := \frac{|r_{k+1}|}{|r_1|}.$$

Hence the method terminates if the residual r_{k+1} becomes sufficiently small in relation to r_1. However, we also want to include a safeguard that stops the algorithm in case the iteration comes to a standstill, i.e., successive iterates cease to change notably. Thus we further introduce

$$(9.4) \quad r_e := \frac{|\boldsymbol{W}_{k+1} - \boldsymbol{W}_k|}{|\boldsymbol{W}_{k+1}|} \quad \text{and} \quad r_e := 0 \text{ if } |\boldsymbol{W}_{k+1}| = 0.$$

and use a combination of r_e and r_b as stopping criterion for our registration strategy.

As briefly addressed in Section 7.3 the developed registration strategy is based on successively deblurring χ_i^{ε} to refine iteratively the computed deformation field. For small values of ε the blurred edge maps χ_i^{ε} approximate the binary edges maps χ_i (since $\chi_i^{\varepsilon} \xrightarrow{\varepsilon \to 0} \chi_i$ pointwise as addressed in Section 8.2) and thus local details in the pathway of the underlying edge sets Γ_i are still clearly visible in χ_i^{ε}. Large values of ε, on the other hand,

---

Algorithm 9.1 Iterative strategy for the elastic registration of edge maps.

    Input: χ_0, χ_1
    Output: $\boldsymbol{W}^*$

1: **Preprocesing:** Center the edge sets Γ_i within Ω
2: **Choose** $k_{\mathrm{inc}} \in \mathbb{N} : k_{\mathrm{inc}} \geq 2$, $\mathrm{tol} \in (0,1)$, $K \in \mathbb{N}$ and $\varepsilon > 0$
3: **Compute** blurred edge maps χ_i^{ε}
4: **Set** $\boldsymbol{W}^* = 0 \in \mathbb{R}^{2N^2}$

5: **for** $\kappa = 1, \ldots, K$ **do**

6:     **Set** $k_{\max} = k_{\mathrm{inc}} \cdot \kappa$, $k = 0$, $r_b = r_e = 2 \cdot \mathrm{tol}$, $\boldsymbol{W}_0 = \boldsymbol{W}^*$ and $\chi_i^{\varepsilon} \leftarrow \left(\chi_i^{\varepsilon}\right)^{\mathbf{2}}$

7:     **while** $\min(r_b, r_e) > \mathrm{tol}$ **and** $k \leq k_{\max}$ **do**

8:         **Solve**
$$\left(-\boldsymbol{E} + \mathfrak{D}\chi_{0,k}^{\varepsilon}\right)\boldsymbol{D}_k = -\boldsymbol{E}\boldsymbol{W}_k - \boldsymbol{F}_k$$

9:         **Compute** t_k according to (9.2)
10:       **Update** $\boldsymbol{W}_{k+1} = \boldsymbol{W}_k + t_k\boldsymbol{D}_k$ and $k \leftarrow k + 1$
11:       **Compute** r_b and r_e given by (9.3) and (9.4) respectively

12:     **end while**

13:     **Update** $\boldsymbol{W}^* = \boldsymbol{W}_k$

14: **end for**

---

introduce excessive blurring in χ_i^{ε} such that small structures in Γ_i may be "washed out". Thus an ill-chosen value for ε may have a negative effect on the registration. Hence the idea is to start with a rather large value of ε to compute a "rough" deformation field reflecting only the most significant deviations of template and reference edge maps. This is realized by choosing the maximal number of iterations $k_{\max}$ for Newton's method particularly small. The computed deformation field serves as an initial guess for the next run of Newton's method for which we decrease the blur in χ_i^{ε} while simultaneously increasing $k_{\max}$. The updated deformation field serves as initial guess for a subsequent run of Newton with further increased $k_{\max}$ and less blurred edge maps and so on. Algorithm 9.1 summarizes this procedure.

Note that we deblur the edge maps χ_i^{ε} by computing their elementwise square, noted $(\chi_i^{\varepsilon})^{\mathbf{2}}$. Since $\chi_i^{\varepsilon}(\boldsymbol{x}) = 1$ on Γ_i and smoothly decreases to zero away from Γ_i, the elementwise squaring has no effect on edge set pixels but lowers the intensity values of pixels around edges. Thus, repeated elementwise squaring successively reduces blur in χ_i^{ε}. Note further that we perform a preprocessing step by centering the edge sets Γ_i within the image domain Ω. This can be seen as a form of rigid preregistration. Since elastic registration is not well suited to account for large translations (see, e.g., the discussion in [Mod04, Sec. 9.1]) this preregistration can reduce the required computational effort considerably.

Note that Algorithm 9.1 was originally designed to register a sequence of edge maps arising from cuts through three-dimensional heart models [FKPP12]. However, numerous practical tests showed that the vanishing diffusion strategy employed in Algorithm 9.1 is a reliable and efficient strategy for the elastic registration of edge maps in general. Thus a natural question is whether a combination of the proposed approach for the computation of edge maps summarized in Algorithm 5.1 and the registration strategy given by Algorithm 9.1 can be employed to register image sequences that are challenging for classical intensity based registration techniques. The next chapter discusses initial results obtained by this combination.

Edge Detection and Registration

In the following we focus on image sequences that manifest pronounced intensity variations not only gradually in space but also abruptly in time, as for instance in a sequence of dynamic contrast enhanced magnetic resonance images (DCE-MRIs). Figure 10.1 shows three members of an artificial image sequence mimicking typical intensity variations seen in MR images of the kidney before, during and after the injection of contrast agent. A common problem in this context is to remove patient motion from the sequence for the purpose of pointwise diagnostic investigation. This task tends to pose a significant challenge for classical intensity based registration strategies due to the substantial intensity variation in the sequence seen before, during and after the appearance of contrast agent. We will show in the following that a combination of Algorithms 5.1 and 9.1 is indeed capable of reliably registering such sequences. All results shown below have been computed on the machine specified in the beginning of Chapter 6. The implementation was again done in MATLAB and Python.

10.1 Combining both Approaches

Before we discuss registration results we address the specifics of this combined strategy. As was the case in Chapter 6, we drop again the distinction between continuum and discrete notation, i.e., we use $\tilde{I}$ instead of $\tilde{I}_h$ or $\tilde{\boldsymbol{I}}_h$ to denote the raw image. However, the objects under consideration are again digital images of size $N \times N$. Thus let $\{\tilde{I}_k\}_{k=1}^M$ denote a sequence of M images. The results shown below were obtained by employing the following algorithmic idea. We use Algorithm 5.1 to compute a sequence of binary edge maps $\{\chi_k\}_{k=1}^M$ associated to $\{\tilde{I}_k\}_{k=1}^M$. Then we apply Algorithm 9.1 to register the sequence $\{1 - \chi_k\}_{k=1}^M$ (since the successive deblurring strategy of Algorithm 9.1 is based on edge maps that have the value one on edges and are zero otherwise). Several important observations should be made at this point. First, note that there are numerous ways to perform this registration. An approach that has already been successfully applied in the context of DCE-MR images is to register consecutive pairs of images in the sequence [Kee10]. Successively registering all pairs in the sequence essentially means that the whole sequence is registered to the first image. However, for standard intensity based approaches a pairwise registration is preferable over registering all images to one single reference since image pairs usually exhibit comparable intensity modulations. The strategy employed here, on the other hand, is based on edge maps and thus its performance is *not* dependent on sufficiently similar intensities of template and reference. Thus we register $\chi_2, \ldots, \chi_M$ to the single reference χ_1.

A further crucial observation is that Algorithm 5.1 computes *binary* edge maps while Algorithm 9.1 is based upon blurred edge maps. Thus a natural question is whether a combination of Algorithm 9.1 with a method that returns fuzzy edge maps would be more
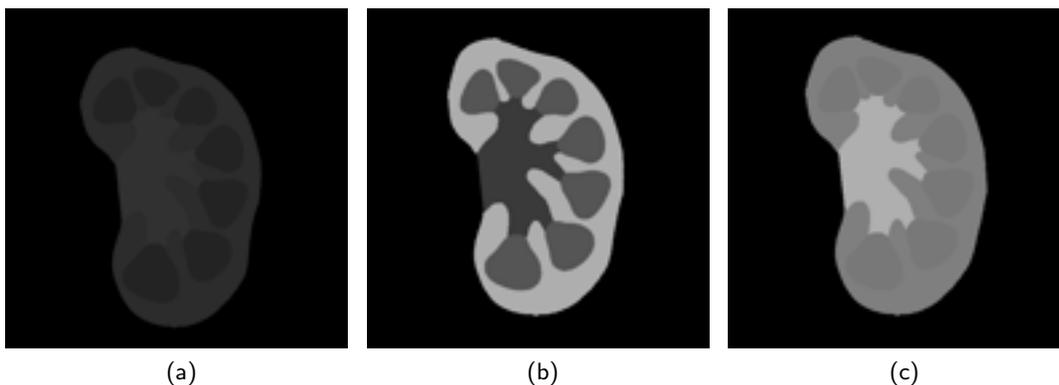
Figure 10.1: Artificial images simulating the effect of contrast agent on the kidneys in a DCE-MRI sequence. Shown is an image before (a), during (b) and after (c) the admission of contrast agent.

Algorithm 10.1 Registration of an image sequence based on edge maps.

Input: $\{\tilde{I}_k\}_{k=1}^M$
Output: $\{\tilde{I}_{k,\mathrm{reg}}\}_{k=1}^M$

1: **Set** $\tilde{I}_{1,\mathrm{reg}} := \tilde{I}_1$
2: **for** $k = 2, \dots M$ **do**
3:      **Compute** χ_k for $\tilde{I}_k$ using Algorithm 5.1
4:      **Compute** $\boldsymbol{w}_k$ by Algorithm 9.1 using $1 - \chi_k$ as template and $1 - \chi_1$ as reference
5:      **Set** $\tilde{I}_{k,\mathrm{reg}} = \tilde{I}_k \circ (\mathrm{id} + \boldsymbol{w}_k)$
6: **end for**

---

appropriate. In fact after completion of the work [FKPP10], numerous practical tests with fuzzy edge maps that were computed by existing approaches have been performed. None of them yielded satisfactory results. For instance, we tried to register inverted, i.e., zero off edges and vice versa, Ambrosio–Tortorelli phase functions using the presented successive deblurring strategy. The main problem of such an approach is that phase functions of non-artificial and possibly noisy images tend to exhibit many weak edges. Thus inverted phase functions are usually very densely supported in the image domain Ω. This can impair the computed deformation field considerably since the registration tries to establish a correspondence between all nonzero regions of the fuzzy edge maps. That usually leads to the unwanted effect that erroneously detected weak edges corresponding to noise are matched to one another. This tends to yield a highly non-linear deformation field.

In contrast, edge maps computed by Algorithm 5.1 have been shown to manifest robustly distinct separated edges that are less likely to be impaired by noise corruption. This serves the successive deblurring strategy of Algorithm 9.1 which proved to perform exceptionally well with binary edge maps computed by our novel approach. Note that we also tested Algorithm 9.1 with Canny edge maps. One of the most severe drawbacks of such an approach is the high sensitivity of Canny's edge detector to the threshold used. While we always employed one set of parameters in Algorithm 5.1 to robustly compute edge maps of the whole image sequence we had to manually adjust the threshold for Canny's detector for each image in the sequence separately. This requirement in addition to the
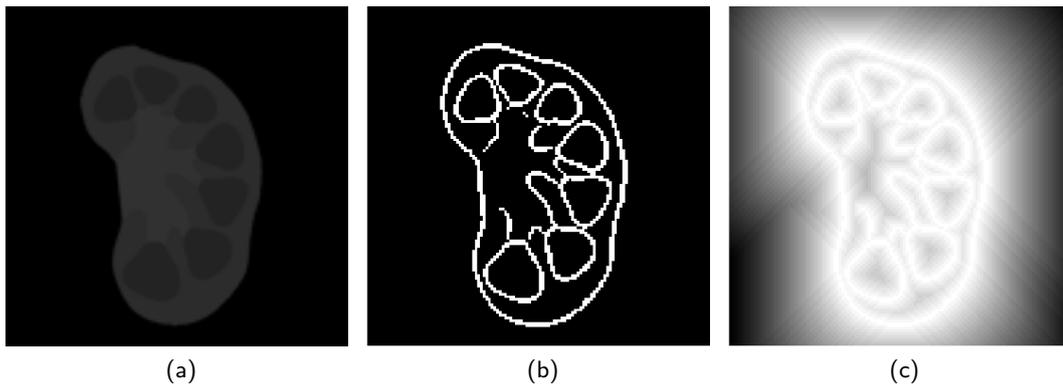
Figure 10.2: First member of the artificial kidney MR-image sequence. Shown is the image (a), the (inverted) edge map computed by Algorithm 5.1 using $\beta = 5$, $\delta = 0.2$ and $\vartheta = 1$ (b) and its blurred version used as reference edge map in Algorithm 9.1.

sensitivity of Canny to noise (see Figure 6.3) and its inability to detect certain types of edges (compare Figure 6.5) made its joint use with Algorithm 9.1 impractical.

Thus we employ Algorithm 5.1 to compute binary edge maps $\{\chi_k\}_{k=1}^M$ of the image sequence $\{\tilde{I}_k\}_{k=1}^M$ to subsequently register $\{1 - \chi_k\}_{k=2}^M$ to $1 - \chi_1$ by Algorithm 9.1. Hence we obtain $M - 1$ deformation fields $\{\boldsymbol{w}_k\}_{k=2}^M$. Since we are interested in a registration of the image sequence $\{\tilde{I}_k\}_{k=1}^M$ we thus apply the computed deformation fields to the images $\{\tilde{I}_k\}_{k=2}^M$. Hence we use the (blurred) edge maps $\{1 - \chi_k\}_{k=2}^M$ in the registration scheme but the registered image sequence is given by $\{\tilde{I}_k \circ (\mathrm{id} + \boldsymbol{w}_k)\}_{k=2}^M$. A heuristic overview of this procedure is given in Algorithm 10.1. In the next section we present first results obtained by using this algorithmic approach.

10.2 Edge Map Based Registration of Image Sequences

For all results in this section the default settings $\mathrm{tol}_{\mathrm{in}} = \mathrm{tol}_{\mathrm{out}} = 1.0e - 2$, $k_{\mathrm{out}}^{\max} = 5$ and $k_{\mathrm{in}}^{\max} = 20$ have been employed in Algorithm 5.1. The registration scheme presented in Algorithm 9.1 always used $\lambda = \mu = 0.1$, $k_{\mathrm{inc}} = 4$, $K = 4$ and $\mathrm{tol} = 1.0e - 3$.

We start by discussing the artificial kidney MR-image sequence introduced in the beginning of this chapter (see Figure 10.1). The sequence consists of $M = 150$ images of resolution 128×128 simulating magnetic resonance recordings of a kidney in the presence of contrast agent. The images do not only manifest intensity variations but also mimic patient motion via translations and deformations of the kidney over time. Note that tissue deformations typically seen in the kidneys during an MR acquisition may be considered as being elastic [HKR$^+$10]. This justifies the use of an elastic penalizer and we thus apply Algorithm 10.1 to register the sequence. Figure 10.2 shows the first image $\tilde{I}_1$, the associated (inverted) edge map computed by Algorithm 5.1 and the initial blurred edge map used in Algorithm 9.1. Note that one set of parameters β, δ and ϑ was employed in Algorithm 5.1 to compute edge maps $\{\chi_k\}_{k=1}^{150}$ for the whole sequence $\{\tilde{I}_k\}_{k=1}^{150}$. Figure 10.3 illustrates the registration scheme by focusing on a representative member of the sequence. The deformation field in Panel (d) shows that the motion within the sequence is indeed non-rigid. Note that the image depicted in Panel (a) manifests a significantly different intensity modulation than the first image of the sequence shown in Panel (a) of Figure 10.2. Nonetheless, using the
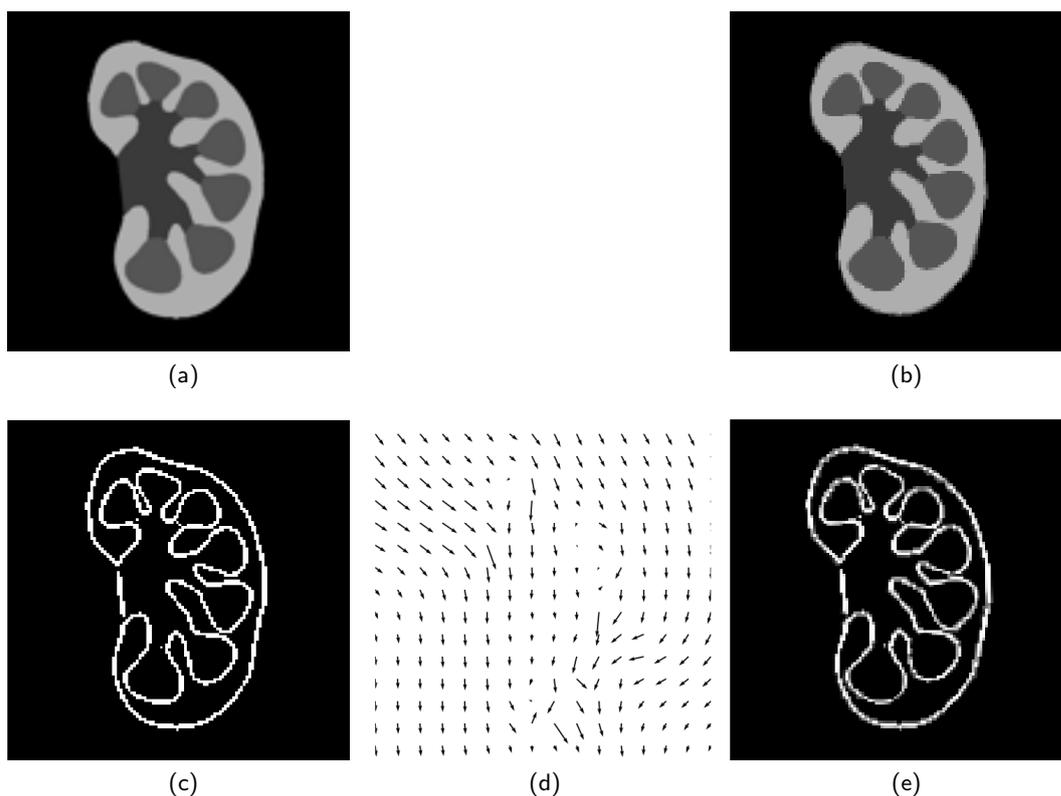
Figure 10.3: Illustration of the edge map based registration approach. The first column shows the image $\tilde{I}_k$ ($k = 23$) and the associated edge map χ_k computed by Algorithm 5.1. Panel (d) depicts the deformation field $\boldsymbol{w}_k$ obtained by using the successive deblurring strategy of Algorithm 9.1. The third column shows the registered image $\tilde{I}_k \circ (\mathrm{id} + \boldsymbol{w}_k)$ and the registered edge map $\chi_k \circ (\mathrm{id} + \boldsymbol{w}_k)$.

same parameters Algorithm 5.1 accurately computes the edge map shown in Panel (c) of Figure 10.3. Panel (e) of the same Figure shows the registered non-blurred template edge map. For comparison, Panel (b) of Figure 10.2 depicts the reference edge map. Visual evaluation of the registered images $\{\tilde{I}_{k,\mathrm{reg}}\}_{k=1}^{150}$ showed that Algorithm 10.1 significantly reduced motion within the sequence.

Next we consider real life images. Recall the microscopic image of cancer cells in a petri dish depicted in Panel (a) of Figure 6.8. The image shown is the first member of a sequence of $M = 180$ raw photomicrographs of a single petri dish acquired at regular time intervals. The objective of the application was to determine division patterns of cancer cells over time. However, the manual placement of the petri dish under the micrograph induced displacements in the sequence. This manual displacement in addition to the natural motility and deformation of cells this made an automatic evaluation of cell divisions intractable. Thus the task was to remove motion to allow for further automatic postprocessing. However, the sequence did not only manifest displacements and cell movement. To differentiate cancerous from healthy cells a chemical agent was employed that caused cancer cells to illuminate spontaneously while healthy cells showed little or no response. This procedure induced large local intensity variations in the images. Hence this sequence posed a registration challenge quite similar to that of the sequence of artificial kidney MR

| (a) | (b) | (c) |
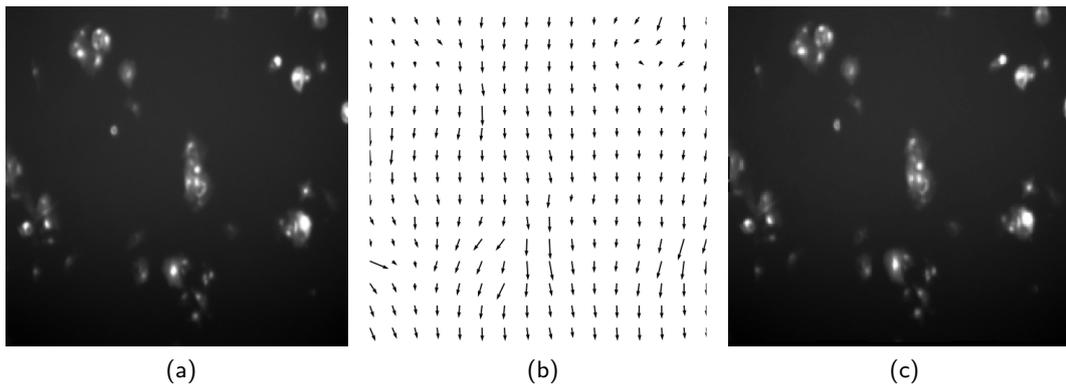
Figure 10.4: Registration of microscopic images based on edge maps. Shown is a member of the sequence (a), the computed deformation field (b) and the registered image (c).

images discussed above. Panel (b) of Figure 6.8 shows the reference edge map χ_1 computed by Algorithm 5.1. Note again that the parameters used to obtain χ_1 were also applied to compute edge maps for the remaining images. Figure 10.4 shows an exemplary member of the sequence and the registration obtained. The computed deformation field indicates that the underlying displacement is indeed not purely rigid. Note that Algorithm 10.1 was able to stabilize the sequence fully. We also tested our registration approach using the more fine-grained edge maps corresponding to $\vartheta = 0.5$ and $\vartheta = 0.25$ (see Figure 6.9). The results obtained were comparable mainly since the topological structure of the edge maps does not change as $\vartheta \to 0$ (compare the discussion in Section 6.4). This stands in stark contrast to edge maps obtained by Canny's edge detector. In Canny's detector thresholding needs to be adjusted not only for each image separately, the computed edge maps may also manifest dramatic structural changes for minimally altered threshold values (compare Figure 6.5). This further highlights the advantages of embedding Algorithm 5.1 in a registration framework.

Finally we return to the sequence of DCE-MRIs of a human torso. A member of this sequence is depicted in Panel (a) of Figure 6.7. The initial image of the sequence whose edge map served as reference in our registration strategy is shown in Figure 2.1. The objective was to remove motion from the sequence for the purpose of pointwise diagnostic investigation. The images manifest pronounced intensity variations not only gradually in space but also abruptly in time. Figure 10.5 illustrates the computed registration by means of an exemplary member of the sequence. Note first that modeling the motion seen in the sequence as subsequent elastic deformations is not new [HKR$^+$10] and can be justified by biomechanical considerations. In contrast to previously published works, our edge map based approach has the advantage of not being dependent on special preprocessing or temporal smoothing techniques. Using Algorithm 10.1 we were indeed able to significantly reduce motion artifacts in the sequence. However, in trying to match all edges in the template edge maps to their corresponding counterparts in the reference edge map, the linearly elastic registration approach induced an unwanted though small deformation in the spinal column (compare the deformation field seen in Panel (b) of Figure 10.5). This violates the strict requirements of clinical diagnostic investigation. A possible remedy for this unwanted effect may lie in the restriction of the linearly elastic penalizer to specific regions in the image. This means to register some areas of the image elastically while others
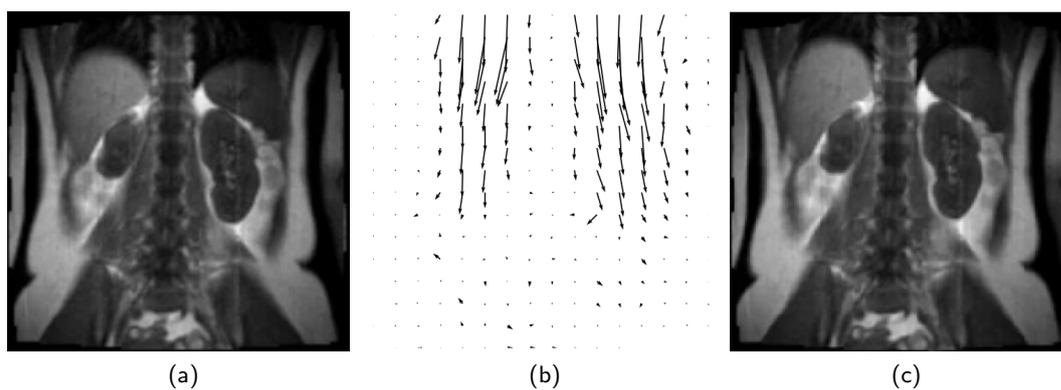
Figure 10.5: Edge map based registration of DCE-MRIs of a human torso. Shown is a member of the sequence (a), the deformation field computed by Algorithm 9.1 (b) and the registered image (c).

are matched according to rigid body motions. Extending Algorithm 9.1 in this manner may have beneficial effects. Similarly, our edge map based registration scheme is of course not limited to a linearly elastic penalty. Depending on the needs of the application at hand, the use of a fluid, diffusion or curvature based regularization (see, e.g., [Mod04]) is straight-forward.

A

Notation

As a rule, albeit not strictly observed, α usually stands for a real number, the letters $i, \ldots, n$ denote integers (following the classic from-"h"-to-"o" water-notation used in FOR-TRAN), where i and j are used for matrix entries, k appears as index in sequences, l is used to indicate subsequences, m denotes the degree of regularity (as in $H^m(\Omega)$ for instance) and n refers to space dimension.

General Notation   Suppose $\boldsymbol{a} = (a_1, \ldots, a_n)$ as well as $\boldsymbol{b} = (b_1, \ldots, b_n)$ are a n-dimensional vectors and $\boldsymbol{A} = (A_{ij})$ as well as $\boldsymbol{B} = (B_{ij})$ are a $m \times n$ matrices.

$\mathbb{N}, \mathbb{R}$ set of natural numbers $\{1, 2, \ldots\}$ and real numbers respectively

$\boldsymbol{x} = (x, y)$ vector (usually in $\Omega \subset \mathbb{R}^2$) with Cartesian coordinates x and y. All vectors, vector valued functions and matrices are written in **bold face**.

$\boldsymbol{a}^\top, \boldsymbol{A}^\top$ transpose of the vector $\boldsymbol{a}$ and transpose of the matrix $\boldsymbol{A}$ respectively

$\det(\boldsymbol{A})$ determinant of the matrix $\boldsymbol{A}$

$\operatorname{tr}(\boldsymbol{A})$ trace of the matrix $\boldsymbol{A}$

$|\alpha|$ absolute value of the real number α

$|\boldsymbol{a}|$ Euclidean norm of the vector $\boldsymbol{a}$, i.e., $|\boldsymbol{a}| = \sqrt{\sum_{i=1}^n a_i^2}$, unless explicitly stated otherwise

$|\boldsymbol{A}|$ Frobenius norm of the matrix $\boldsymbol{A}$, i.e., $|\boldsymbol{A}| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$

$|\omega|$ Lebesgue measure of the set ω

$\boldsymbol{a} \cdot \boldsymbol{b}$ usual scalar product of the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e., $\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^n a_i b_i$

$\boldsymbol{A} : \boldsymbol{B}$ Frobenius inner product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, i.e.,

$$\boldsymbol{A} : \boldsymbol{B} = \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ij} = \operatorname{tr}(\boldsymbol{A}^\top \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}\boldsymbol{B}^\top).$$

Note that this component-wise inner product induces the Frobenius norm, i.e., $|\boldsymbol{A}| = \sqrt{\boldsymbol{A} : \boldsymbol{A}}$.

$\partial\omega$ boundary of the set ω

$\bar{\omega}$ closure of the set ω, i.e., $\bar{\omega} = \omega \cup \partial\omega$

$\mathfrak{X} := \mathfrak{Y}$ expression $\mathfrak{Y}$ defines expression $\mathfrak{X}$

a.e.   abbreviation for almost everywhere, i.e., some property holds everywhere except on a set of Lebesgue measure zero.

Functions and Derivatives   Let $d, n \in \mathbb{N}$, $\Omega \subset \mathbb{R}^n$ be open, bounded and non-empty, $\boldsymbol{x} = (x_1, \ldots, x_n)$ be a vector and $u : \Omega \to \mathbb{R}$ as well as $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ $(d > 1)$ be functions.

$\mathrm{Rg}(u)$ range of the function u, i.e., $\mathrm{Rg}(u) = \{u(\boldsymbol{x}) \in \mathbb{R} \,|\, \boldsymbol{x} \in \Omega\}$

$\mathrm{supp}(u)$ support of u, i.e., $\mathrm{supp}(u) = \{\boldsymbol{x} \in \Omega \,|\, u(\boldsymbol{x}) \neq 0\}$

$\ker(u)$ kernel of u, i.e., $\ker(u) = \{\boldsymbol{x} \in \Omega \,|\, u(\boldsymbol{x}) = 0\}$

id identity function, i.e., $\mathrm{id}(\boldsymbol{x}) = \boldsymbol{x}$

$\frac{\partial}{\partial x_i} u(\boldsymbol{x})$ i-th partial derivative of u, i.e., $\frac{\partial}{\partial x_i} u(\boldsymbol{x}) := \lim_{h \to 0} \frac{u(\boldsymbol{x} + h\boldsymbol{e}_i) - u(\boldsymbol{x})}{h}$ (if this limit exists) where $\boldsymbol{e}_i$ denotes the i-th canonical basis vector in $\mathbb{R}^n$. Analogously we define $\frac{\partial^2}{\partial x_i \partial x_j} u(\boldsymbol{x}) := \frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_j} u(\boldsymbol{x}) \right)$ and so on. We employ multi-indices to introduce an efficient shorthand representation for higher order partial derivatives. Thus let $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ be a n-tuple of natural numbers. Then let

$$D^{\boldsymbol{\alpha}} u(\boldsymbol{x}) = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}} u(\boldsymbol{x}),$$

denote the partial derivative of u of order $|\boldsymbol{\alpha}| := \sum_{i=1}^n \alpha_i$.

∇u gradient of u, i.e., $\nabla u := (\frac{\partial}{\partial x_1} u(\boldsymbol{x}), \frac{\partial}{\partial x_2} u(\boldsymbol{x}), \ldots, \frac{\partial}{\partial x_n} u(\boldsymbol{x}))^\top \in \mathbb{R}^n$. Note that we always regard ∇u as column vector.

$\nabla \boldsymbol{u}$ Jacobian matrix of the vector field $\boldsymbol{u}$, i.e.,

$$\nabla \boldsymbol{u} := \begin{pmatrix} \frac{\partial}{\partial x_1} u_1(\boldsymbol{x}) & \frac{\partial}{\partial x_2} u_1(\boldsymbol{x}) & \cdots & \frac{\partial}{\partial x_n} u_1(\boldsymbol{x}) \\ \frac{\partial}{\partial x_1} u_2(\boldsymbol{x}) & \frac{\partial}{\partial x_2} u_2(\boldsymbol{x}) & \cdots & \frac{\partial}{\partial x_n} u_2(\boldsymbol{x}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial}{\partial x_1} u_d(\boldsymbol{x}) & \frac{\partial}{\partial x_2} u_d(\boldsymbol{x}) & \cdots & \frac{\partial}{\partial x_n} u_d(\boldsymbol{x}) \end{pmatrix} \in \mathbb{R}^{d \times n}$$

$\nabla^2 u$ Hessian matrix of u, i.e.,

$$\nabla^2 u := \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} u(\boldsymbol{x}) & \frac{\partial^2}{\partial x_1 \partial x_2} u(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} u(\boldsymbol{x}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} u(\boldsymbol{x}) & \frac{\partial^2}{\partial x_2^2} u(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_n} u(\boldsymbol{x}) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} u(\boldsymbol{x}) & \frac{\partial^2}{\partial x_n \partial x_2} u(\boldsymbol{x}) & \cdots & \frac{\partial^2}{\partial x_n^2} u(\boldsymbol{x}) \end{pmatrix} \in \mathbb{R}^{n \times n}$$

$\nabla \cdot \boldsymbol{u}$ divergence of the vector field $\boldsymbol{u}$, i.e.,

$$\nabla \cdot \boldsymbol{u} := \frac{\partial}{\partial x_1} u_1(\boldsymbol{x}) + \frac{\partial}{\partial x_2} u_2(\boldsymbol{x}) + \cdots + \frac{\partial}{\partial x_n} u_n(\boldsymbol{x}) = \sum_{i=1}^n \frac{\partial}{\partial x_i} u_i(\boldsymbol{x})$$

Δu Laplacian of u, i.e.,

$$\Delta u := \frac{\partial^2}{\partial x_1^2} u(\boldsymbol{x}) + \cdots + \frac{\partial^2}{\partial x_n^2} u(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} u(\boldsymbol{x}) = \mathrm{tr}(\nabla^2 u) = \nabla \cdot \nabla u$$

Function Spaces   Note that we understand the following spaces to be naturally extended to vector fields, e.g., $\boldsymbol{u} \in C(\Omega)$ means $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ is continuous (where the dimension $d > 1$ is always explicitly stated in the text).

$C(\Omega)$ space of continuous functions from Ω to $\mathbb{R}$

$C^m(\Omega)$ space of m-times ($m \in \mathbb{N}$) continuously differentiable functions on Ω, i.e.,

$$C^m(\Omega) = \{u \in C(\Omega) \,|\, D^{\boldsymbol{\alpha}} u \in C(\Omega), \forall \boldsymbol{\alpha} : |\boldsymbol{\alpha}| \leq m \}$$

$C^\infty(\Omega)$ space of infinitely differentiable functions on Ω, i.e., $C^\infty(\Omega) = \bigcap_{m=0}^{\infty} C^m(\Omega)$ where $C^0(\Omega) := C(\Omega)$

$C_0^m(\Omega)$ subspace of functions in $C^m(\Omega)$ with support compactly contained in Ω ($m = 0, 1, \ldots, \infty$)

$L^p(\Omega)$ space of p-integrable ($1 \leq p < \infty$) or essentially bounded ($p = \infty$) functions on Ω, i.e.,

$$L^p(\Omega) = \left\{ u : \Omega \to \mathbb{R} \,\middle|\, u \text{ is Lebesgue measurable and } \|u\|_{L^p(\Omega)} < \infty \right\},$$

with

$$\|u\|_{L^p(\Omega)} = \begin{cases} \left(\int_\Omega |u|^p \, d\boldsymbol{x} \right)^{\frac{1}{p}}, & 1 \leq p < \infty, \\ \operatorname{ess\,sup}_{\boldsymbol{x} \in \Omega} |u(\boldsymbol{x})|, & p = \infty, \end{cases}$$

where

$$\operatorname{ess\,sup}_{\boldsymbol{x} \in \Omega} |u(\boldsymbol{x})| = \inf \left\{ \xi \in \mathbb{R} \,\middle|\, |\{\boldsymbol{x} \in \Omega \,|\, u(\boldsymbol{x}) > \xi \}| > 0 \right\}.$$

$W^{m,p}(\Omega)$ Sobolev space of L^p-functions for which weak partial derivatives of order m exist and belong to $L^p(\Omega)$, i.e.,

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega) \,|\, D^{\boldsymbol{\alpha}} u \in L^p(\Omega), \forall \boldsymbol{\alpha} : |\boldsymbol{\alpha}| \leq m \}.$$

We consider these spaces to be equipped with the usual Sobolev norm

$$\|u\|_{W^{m,p}(\Omega)} = \begin{cases} \left(\sum_{|\boldsymbol{\alpha}| \leq k} \|D^{\boldsymbol{\alpha}} u\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}}, & 1 \leq p < \infty, \\ \sum_{|\boldsymbol{\alpha}| \leq k} \|D^{\boldsymbol{\alpha}} u\|_{L^\infty(\Omega)}, & p = \infty. \end{cases}$$

$H^m(\Omega)$ Hilbert space of square integrable Sobolev functions, i.e., $H^m(\Omega) = W^{m,2}(\Omega)$. In this work mainly the spaces $H^1(\Omega)$ and $H^2(\Omega)$ are used. For these spaces the associated norms can be written in terms of the previously introduced notation

$$\|u\|_{H^1(\Omega)}^2 = \|u\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2,$$

and

$$\|u\|_{H^2(\Omega)}^2 = \|u\|_{L^2(\Omega)}^2 + \|\nabla u\|_{L^2(\Omega)}^2 + \left\|\nabla^2 u\right\|_{L^2(\Omega)}^2,$$

where $\|\nabla u\|_{L^2(\Omega)} := \||\nabla u|\|_{L^2(\Omega)}$ and $\left\|\nabla^2 u\right\|_{L^2(\Omega)} := \left\||\nabla^2 u|\right\|_{L^2(\Omega)}$ with $|\cdot|$ as given above.

Topology and Convergence   Let $\{f_k\}_{k\geq 1}$ and f be measurable functions mapping from Ω to $\mathbb{R}^d$. Further, let X, Y be real normed spaces with norms $|\cdot|_X$ and $|\cdot|_Y$ respectively, let $u \in X$ and $\{u_k\}_{k\geq 1}$ be a sequence in X.

$\mathcal{L}(X, Y)$ space of continuous linear operators from X to Y, i.e.,

$$\mathcal{L}(X, Y) = \left\{ T : X \to Y \,\middle|\, T \text{ is linear and } \|T\|_{X,Y} < \infty \right\},$$

with

$$\|T\|_{X,Y} = \sup \left\{ |Tu|_Y \,\middle|\, u \in X : |u|_X \leq 1 \right\}$$

X' dual space of X, i.e., $X' = \mathcal{L}(X, \mathbb{R})$

$u_k \xrightarrow[X]{} u$ $\{u_k\}_{k\geq 1}$ converges to u in the strong topology on X, i.e., $\lim_{k\to\infty} |u_k - u|_X = 0$

$u_k \underset{X}{\rightharpoonup} u$ $\{u_k\}_{k\geq 1}$ converges to u in the weak topology on X, i.e., $\forall l \in X' : \lim_{k\to\infty} l[u_k] = l[u]$

$f_k \xrightarrow{\text{a.e.}} f$ $\{f_k\}_{k\geq 1}$ converges to f almost everywhere on Ω, i.e., $\lim_{k\to\infty} f_k(\boldsymbol{x}) = f(\boldsymbol{x})$ a.e. Ω

Calculus Facts

In the following we summarize classical results that are used frequently in this work. Note that this synopsis is by no means exhaustive and just outlines important facts. For the statements presented below suppose that $d, n \in \mathbb{N}$, let $\Omega \subset \mathbb{R}^n$ be open, bounded and non-empty, let X be a real Banach space with norm $|\cdot|_X$ and let H be a real Hilbert space with norm $|\cdot|_H$.

B.1 Inequalities

We repeatedly employ the following estimates.

Chebyshev's Inequality   See, e.g., [Ash72, Sec. 2.4.9]. Let $f : \Omega \to \mathbb{R}$ be measurable. Then for all $0 < p < \infty$ and any $\varepsilon > 0$

$$|\{\boldsymbol{x} \in \Omega \,|\, |f(\boldsymbol{x})| \geq \varepsilon \}| \leq \frac{1}{\varepsilon^p} \int_\Omega |f|^p \, d\boldsymbol{x}.$$

Note that the cases $p = 1$ and $p = 2$ are of particular interest in this work. For $p = 1$ this estimate is often referred to as the *Markov inequality*, the case $p = 2$ is commonly known as *Chebyshev's inequality*.

Cauchy's Inequality   See, e.g., [Eva08, App. B.2]. Let $\alpha, \beta \in \mathbb{R}$. Then $\alpha\beta \leq \frac{\alpha^2}{2} + \frac{\beta^2}{2}$ and if $\alpha, \beta \geq 0$ we have for any $\varepsilon > 0$ that $\alpha\beta \leq \varepsilon\alpha^2 + \frac{\beta^2}{4\varepsilon}$.

Cauchy–Schwarz Inequality   See, e.g., [Ada75, Cor. 2.12]. Let $u, v \in L^2(\Omega)$ then

$$|\langle u, v \rangle_{L^2(\Omega)}| \leq \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)}.$$

.

Young's Inequality for Convolutions   See, e.g., [RS75, Sec. IX.4]. Let $1 \leq p, q, r \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$. Then for $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$ we have

$$\|f * g\|_{L^r(\Omega)} \leq \|f\|_{L^p(\Omega)} \|g\|_{L^q(\Omega)}.$$

B.2 Analytical Facts

We often rely on the classical facts given below.

Gâteaux Derivative See, e.g., [AK06, Def. 2.1.5]. Let $F : X \to \mathbb{R}$ and $u \in X$. If the limit

$$\frac{\delta F}{\delta u}[u; v] := \lim_{s \to 0} \frac{1}{s} \left(F[u + sv] - F[u] \right) = \frac{d}{ds} F[u + sv] \bigg|_{s=0},$$

exists for some $v \in X$ it is called the *directional derivative* of F at u in direction v. If the limit exists for all $v \in X$ then F is said to be *Gâteaux differentiable at u* with *Gâteaux derivative* $\frac{\delta F}{\delta u}$.

The Symmetric Difference of Sets See, e.g., [Hal74, Chap. 1 §3]. Let $(\mathcal{B}(\Omega), |\cdot|)$ be the usual Borel–Lebesgue measure space, i.e., $\mathcal{B}(\Omega)$ is the Borel σ-algebra over Ω and $|\cdot|$ denotes the Lebesgue measure. For two sets $A, B \in \mathcal{B}(\Omega)$ we define *the symmetric difference of two sets* by

$$A \triangle B := (A \backslash B) \cup (B \backslash A),$$

or equivalently

$$A \triangle B = (A \cup B) \backslash (A \cap B).$$

The symmetric difference has the following properties:

- $A \triangle B = B \triangle A$ (symmetry)

- $(A \triangle B) \triangle C = A \triangle (B \triangle C)$ (associativity)

- $A \triangle \emptyset = A$

- $A \triangle A = \emptyset$

- $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$

The Metric Space $\mathcal{M}(\Omega)$ See, e.g., [Hal74, p. 168]. The symmetric difference can be used to turn the measure space $(\mathcal{B}(\Omega), |\cdot|)$ into a metric space. We introduce the mapping $d : \mathcal{B}(\Omega) \times \mathcal{B}(\Omega) \to \mathbb{R} \cup \{\infty\}$ defined by $d(A, B) = |A \triangle B|$. Then d is clearly symmetric and non-negative. Further,

$$(A \triangle C) \triangle (C \triangle B) = A \triangle B,$$

which yields

$$A \triangle B \subset ((A \triangle C) \cup (C \triangle B)),$$

and thus implies the triangle inequality

$$|A \triangle B| \leq |(A \triangle C) \cup (C \triangle B)| \leq |A \triangle C| + |C \triangle B|.$$

However, for zero measure sets, $d(A, B) = 0$ even if $A \neq B$ which makes $(\mathcal{B}(\Omega), d)$ a pseudo-metric space. Thus we introduce $\mathcal{M}(\Omega)$ as the quotient space of $(\mathcal{B}(\Omega), d)$ with respect to the kernel of d, i.e., $\mathcal{M}(\Omega) = (\mathcal{B}(\Omega), d) / \ker(d)$. Then $\mathcal{M}(\Omega)$ is a metric space. Note that two sets in $\mathcal{M}(\Omega)$ are hence identified if they differ only by a set of Lebesgue measure zero.

Schauder's Fixed Point Theorem   See, e.g., [Sch30]. Let K be a convex subset of X and let $\Phi : K \to X$ be continuous. If $\Phi(K)$ is compact in K then Φ has a fixed point in K.

The Lax–Milgram Lemma   See, e.g., [Eva08, Sec. 6.3.1]. Let $a : H \times H \to R$ be a bilinear form. Assume a is bounded, i.e., there exists a positive constant α_{b} such that

$$|a[u,v]| \leq \alpha_{\mathrm{b}} \, |u|_H \, |v|_H \,, \quad \forall u, v \in H.$$

Suppose further that a is coercive, i.e., there exists $\alpha_{\mathrm{c}} > 0$ such that

$$a[u,u] \geq \alpha_{\mathrm{c}} \, |u|_H^2 \,, \quad \forall u \in H,$$

or equivalently

$$\alpha_{\mathrm{c}} = \inf_{u \in H} \frac{a[u,u]}{|u|_H^2} > 0.$$

Let $f : H \to \mathbb{R}$ be a bounded linear functional on H, i.e., there exists $\alpha_f > 0$ such that $|f[v]| \leq \alpha_f \, |v|_H$ for all $v \in H$. Then there exists a unique element $u^* \in H$ such that

$$a[u^*, v] = f[v], \quad \forall v \in H.$$

Weak Sequential Compactness   See, e.g., [AK06, Sec. 2.1.1]. Let the space X be reflexive, i.e., $(X')' = X$. Suppose the sequence $\{x_k\}_{k \geq 1} \subset X$ is bounded in X, i.e., for all $k \in \mathbb{N}$: $|x_k| \leq c$ for some $c > 0$. Then $\{x_k\}_{k \geq 1}$ has a weakly convergent subsequence $\{x_{k_l}\}_{l \geq 1}$ which means there exists $x^* \in X$ such that $x_{k_l} \underset{X}{\rightharpoonup} x^*$ as $l \to \infty$.

C

Addenda

The following items have been left out in the main part mostly for the sake of readability and to avoid cluttering up the text. However, the reader may be interested in some of the details presented below.

C.1 Euler–Lagrange Equation for J_{TA}

Recall that

$$J_{\mathrm{TA}}[I] := \frac{1}{2} \int_\Omega \frac{1}{\nu} \left| I - \tilde{I} \right|^2 + |\nabla I|^2 \, d\boldsymbol{x}.$$

Note that J_{TA} is everywhere Gâteaux differentiable. Thus we start by computing the Gâteaux derivative of J_{TA} in an arbitrary direction $v \in C^\infty(\bar{\Omega})$. Using partial integration we obtain

$$\frac{\delta J_{\mathrm{TA}}}{\delta I}[I;v] = \frac{d}{ds} \left(\frac{1}{2} \int_\Omega \frac{1}{\nu} \left| I + sv - \tilde{I} \right|^2 + |\nabla(I + sv)|^2 \, d\boldsymbol{x} \right) \bigg|_{s=0}$$

$$= \int_\Omega \frac{1}{\nu}(I - \tilde{I})v + \nabla I \cdot \nabla v \, d\boldsymbol{x}$$

$$= \int_\Omega \frac{1}{\nu}(I - \tilde{I})v + \int_{\partial\Omega} v \nabla I \cdot \boldsymbol{n} \, dS - \int_\Omega v \Delta u \, d\boldsymbol{x},$$

where $\boldsymbol{n}$ denotes the outer unit normal vector on $\partial\Omega$. The weak necessary optimality condition for the minimization of J_{TA} is (compare, e.g., [Lue69, Sec. 7.4, Th. 1])

$$\frac{\delta J_{\mathrm{TA}}}{\delta I}[I;v] = 0, \quad \forall v \in C^\infty(\bar{\Omega}).$$

Under the assumption that I is sufficiently regular we may apply the fundamental Lemma of calculus of variations [Ada75, Cor. 3.26] to obtain the Euler–Lagrange equation

$$\begin{cases} -\nu \Delta I + I = \tilde{I}, & \text{in } \Omega, \\ \dfrac{\partial}{\partial \boldsymbol{n}} I = 0, & \text{on } \partial\Omega. \end{cases}$$

C.2 Stencils for Matrices discussed in Section 5.3

We illustrate the nonzero structure of the involved matrices by giving explicit representations of stencil weights for neighbors of field cells. In the case of boundary cells the stencils are modified by adequate zero padding of χ_h outside of the grid.

We start by considering the bending matrix $\boldsymbol{A}[\chi_h]$ associated to A_h given by (5.5). Note that $S_h^{(2)}(\Omega)$ was chosen to be the approximation space in (5.4). Hence, to respect the support of quadratic splines (compare Figure 5.1) we have to consider the 3×3 neighborhood of a field cell. Recall that $\chi_h \in S_h^{(0)}(\Omega)$ and thus χ_h is cellwise constant. The values of χ_h are lexicographically ordered. Thus, to simplify notation suppose χ_h has the following values on a prototype field cell

$$
\begin{array}{ccc}
\xi_7 & \xi_8 & \xi_9 \\
\xi_4 & \xi_5 & \xi_6 \\
\xi_1 & \xi_2 & \xi_3
\end{array}
$$

Then the nonzero structure of $\boldsymbol{A}[\chi_h]$ is based on the following stencil

$$
\frac{1}{360h^4}
\begin{pmatrix}
a_{11}^\star & a_{12}^\star & a_{13}^\star & a_{14}^\star & a_{15}^\star \\
a_{21}^\star & a_{22}^\star & a_{23}^\star & a_{24}^\star & a_{25}^\star \\
a_{31}^\star & a_{32}^\star & a_{33}^\star & a_{34}^\star & a_{35}^\star \\
a_{41}^\star & a_{42}^\star & a_{43}^\star & a_{44}^\star & a_{45}^\star \\
a_{51}^\star & a_{52}^\star & a_{53}^\star & a_{54}^\star & a_{55}^\star
\end{pmatrix},
$$

where

$$
\begin{aligned}
a_{11}^\star &= 26\xi_7, & a_{21}^\star &= 53(\xi_7 + \xi_8), \\
a_{12}^\star &= 53(\xi_4 + \xi_7), & a_{22}^\star &= -136(\xi_4 + \xi_5 + \xi_7 + \xi_8), \\
a_{13}^\star &= -19\xi_1 + 134\xi_4 - 19\xi_7, & a_{23}^\star &= -37\xi_1 - 37\xi_2 - 208\xi_4 - 208\xi_5 - 37\xi_7 - 37\xi_8, \\
a_{14}^\star &= 53(\xi_1 + \xi_4), & a_{24}^\star &= -136(\xi_1 + \xi_2 + \xi_4 + \xi_5), \\
a_{15}^\star &= 26\xi_1, & a_{25}^\star &= 53(\xi_1 + \xi_2),
\end{aligned}
$$

$$
\begin{aligned}
a_{31}^\star &= -19\xi_7 + 134\xi_8 - 19\xi_9, \\
a_{32}^\star &= -37\xi_4 - 208\xi_5 - 37\xi_6 - 37\xi_7 - 208\xi_8 - 37\xi_9, \\
a_{33}^\star &= 2(58\xi_1 + 157\xi_2 + 58\xi_3 + 157\xi_4 + 688\xi_5 + 157\xi_6 + 58\xi_7 + 157\xi_8 + 58\xi_9), \\
a_{34}^\star &= -37\xi_1 - 208\xi_2 - 37\xi_3 - 37\xi_4 - 208\xi_5 - 37\xi_6, \\
a_{35}^\star &= -19\xi_1 + 134\xi_2 - 19\xi_3,
\end{aligned}
$$

$$
\begin{aligned}
a_{41}^\star &= 53(\xi_8 + \xi_9), & a_{51}^\star &= 26\xi_9, \\
a_{42}^\star &= -136(\xi_5 + \xi_6 + \xi_8 + \xi_9), & a_{52}^\star &= 53(\xi_6 + \xi_9), \\
a_{43}^\star &= -37\xi_2 - 37\xi_3 - 208\xi_5 - 208\xi_6 - 37\xi_8 - 37\xi_9, & a_{53}^\star &= -19\xi_3 + 134\xi_6 - 19\xi_9, \\
a_{44}^\star &= -136(\xi_2 + \xi_3 + \xi_5 + \xi_6), & a_{54}^\star &= 53(\xi_3 + \xi_6), \\
a_{45}^\star &= 53(\xi_2 + \xi_3), & a_{55}^\star &= 26\xi_3.
\end{aligned}
$$

Using the same format the following stencil constructs the Gram matrix $\boldsymbol{G}[\chi_h]$ of A_h

$$\frac{1}{14\,400} \begin{pmatrix} g^\star_{11} & g^\star_{12} & g^\star_{13} & g^\star_{14} & g^\star_{15} \\ g^\star_{21} & g^\star_{22} & g^\star_{23} & g^\star_{24} & g^\star_{25} \\ g^\star_{31} & g^\star_{32} & g^\star_{33} & g^\star_{34} & g^\star_{35} \\ g^\star_{41} & g^\star_{42} & g^\star_{43} & g^\star_{44} & g^\star_{45} \\ g^\star_{51} & g^\star_{52} & g^\star_{53} & g^\star_{54} & g^\star_{55} \end{pmatrix},$$

where

$$
\begin{aligned}
g^\star_{11} &= \xi_7, & g^\star_{21} &= 13(\xi_7 + \xi_8), \\
g^\star_{12} &= 13(\xi_4 + \xi_7), & g^\star_{22} &= 169(\xi_4 + \xi_5 + \xi_7 + \xi_8), \\
g^\star_{13} &= 6(\xi_1 + 9\xi_4 + \xi_7), & g^\star_{23} &= 78(\xi_1 + \xi_2 + 9\xi_4 + 9\xi_5 + \xi_7 + \xi_8), \\
g^\star_{14} &= 13(\xi_1 + \xi_4), & g^\star_{24} &= 169(\xi_1 + \xi_2 + \xi_4 + \xi_5), \\
g^\star_{15} &= \xi_1, & g^\star_{25} &= 13(\xi_1 + \xi_2),
\end{aligned}
$$

$$
\begin{aligned}
g^\star_{31} &= 6(\xi_7 + 9\xi_8 + \xi_9), \\
g^\star_{32} &= 78(\xi_4 + 9\xi_5 + \xi_6 + \xi_7 + 9\xi_8 + \xi_9), \\
g^\star_{33} &= 36(\xi_1 + 9\xi_2 + \xi_3 + 9\xi_4 + 81\xi_5 + 9\xi_6 + \xi_7 + 9\xi_8 + \xi_9), \\
g^\star_{34} &= 78(\xi_1 + 9\xi_2 + \xi_3 + \xi_4 + 9\xi_5 + \xi_6), \\
g^\star_{35} &= 6(\xi_1 + 9\xi_2 + \xi_3),
\end{aligned}
$$

$$
\begin{aligned}
g^\star_{41} &= 13(\xi_8 + \xi_9), & g^\star_{51} &= \xi_9, \\
g^\star_{42} &= 169(\xi_5 + \xi_6 + \xi_8 + \xi_9), & g^\star_{52} &= 13(\xi_6 + \xi_9), \\
g^\star_{43} &= 78(\xi_2 + \xi_3 + 9\xi_5 + 9\xi_6 + \xi_8 + \xi_9), & g^\star_{53} &= 6(\xi_3 + 9\xi_6 + \xi_9), \\
g^\star_{44} &= 169(\xi_2 + \xi_3 + \xi_5 + \xi_6), & g^\star_{54} &= 13(\xi_3 + \xi_6), \\
g^\star_{45} &= 13(\xi_2 + \xi_3), & g^\star_{55} &= \xi_3.
\end{aligned}
$$

Since $S_h^{(1)}(\Omega)$ was selected as approximation space in (5.11) the stencil for the stiffness and Gram matrix associated to $B_{\mathrm{b},h}$ defined by (5.12) is smaller then the one for $\boldsymbol{A}[\chi_h]$ and $\boldsymbol{G}[\chi_h]$ respectively. Since linear splines have smaller support than quadratic splines (compare again Figure 5.1) we only have to consider the following four values of χ_h on a 2×2 prototype field cell

$$
\begin{matrix}
\xi_3 & \xi_4 \\
\xi_1 & \xi_2
\end{matrix}
$$

Then the nonzero structure of the stiffness matrix $\boldsymbol{B}[\chi_h]$ is based on the stencil

$$\frac{1}{6h^2} \begin{pmatrix} -2\xi_3 & -\xi_1 - \xi_3 & 2\xi_1 \\ -\xi_3 - \xi_4 & 4(\xi_1 + \xi_2 + \xi_3 + \xi_4) & -\xi_1 - \xi_2 \\ -2\xi_4 & -\xi_2 - \xi_4 & -2\xi_2 \end{pmatrix}.$$

The Gram matrix $\boldsymbol{C}[\chi_h]$ for $B_{\mathrm{b},h}$ is constructed using the stencil

$$\frac{1}{36}\begin{pmatrix} \xi_3 & 2(\xi_1+\xi_3) & \xi_1 \\ 2(\xi_3+\xi_4) & 4(\xi_1+\xi_2+\xi_3+\xi_4) & 2(\xi_1+\xi_2) \\ \xi_4 & 2(\xi_2+\xi_4) & \xi_2 \end{pmatrix}.$$

Stencils of $\boldsymbol{B}[1-\chi_h]$ and $\boldsymbol{C}[1-\chi_h]$ associated to $B_{\mathrm{f},h}$ given by (5.14) are obtained by replacing ξ_i with $1-\xi_i$ in their respective counterparts. Finally, the projection matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are obtained by using the stencils

$$\frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

and

$$\frac{1}{36}\begin{pmatrix} 1 & 4 & 1 \\ 4 & 16 & 4 \\ 1 & 4 & 1 \end{pmatrix},$$

respectively.

C.3 Stencils for the Discrete Elasticity Operator

We illustrate the nonzero structure of the matrices $\boldsymbol{E}^{i,j}$ by explaining the discretization of the associated operators $\mathcal{E}^{i,j}$, $1 \leq i,j \leq 2$, using explicit representations of stencil weights for cells near a corner of the grid Ω_h.
We start by considering the matrix $\boldsymbol{E}^{1,1}$. The following stencils discretize the operator $\mathcal{E}^{1,1}$ near the lower left corner of the grid

$$(\mathrm{C}.1) \quad \frac{1}{4h^2} \begin{pmatrix} 0 & 0 & -2\mu \\ 0 & 8\mu+4\lambda & -4\mu-4\lambda \\ 0 & 0 & -2\mu \\ 0 & 0 & -2\mu \\ 0 & 2\lambda+4\mu & -2\mu-2\lambda \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -2\mu & 0 & -2\mu \\ -4\mu-4\lambda & 16\mu+8\lambda & -4\mu-4\lambda \\ -2\mu & 0 & -2\mu \\ -2\mu & 0 & -2\mu \\ -2\mu-2\lambda & 8\mu+4\lambda & -2\mu-2\lambda \\ 0 & 0 & 0 \end{pmatrix},$$

where the upper right block holds the stencil weights for neighbors of a field cell while the other blocks show stencil weights for neighbors of boundary cells. Using the same format the stencils discretizing the operator $\mathcal{E}^{1,2}$ near the same corner of Ω_h are given by

$$(\mathrm{C}.2) \quad \frac{1}{4h^2} \begin{pmatrix} 0 & \mu-\lambda & -\mu-\lambda \\ 0 & 0 & 0 \\ 0 & -\mu-\lambda & \mu+\lambda \\ 0 & \mu-\lambda & -\mu-\lambda \\ 0 & \mu+\lambda & -\mu+\lambda \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu+\lambda & 0 & -\mu-\lambda \\ 0 & 0 & 0 \\ -\mu-\lambda & 0 & \mu+\lambda \\ \mu+\lambda & 0 & -\mu-\lambda \\ \mu-\lambda & 0 & -\mu-\lambda \\ 0 & 0 & 0 \end{pmatrix},$$

which illustrates the nonzero structure of the matrix $\boldsymbol{E}^{1,2}$. The stencils for $\boldsymbol{E}^{2,2}$ and $\boldsymbol{E}^{2,1}$ are obtained by adequately copying and mirroring (C.1) and (C.2) respectively.

Bibliography

[Ada75]   R.A. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.

[AK06]   G. Aubert and P. Kornprobst. *Mathematical Problems In Image Processing*, volume 147 of *Applied Mathematical Sciences*. Springer, 2nd edition, 2006.

[Ash72]   R. B. Ash. *Real Analysis and Probability*. Academic Press, New York, 1972.

[AST12]   I. Abu-Falahah, P. R. Stinga, and J. L. Torrea. A note on the almost every-where convergence to initial data for some evolution equations. *ArXiv e-prints*, June 2012.

[AT90]   L. Ambrosio and V.M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Communications on Pure and Applied Mathematics*, 43:999–1036, 1990.

[Bis06]   C.M. Bishop. *Pattern recognition and machine learning*. Information science and statistics. Springer, 2006.

[BKP10]   K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[BL10]   J.M. Borwein and R.L. Luke. *Handbook of Mathematical Methods in Imaging*, chapter Duality and Convex Programming, pages 229–270. Springer, 2010.

[Bre04]   S. C. Brenner. Korn's inequalities for piecewise H^1 vector fields. *Mathematics of Computation*, 73:1067–1087, 2004.

[Byu98]   Du-Won Byun. An inequality on solutions of heat equation. *Journal of Inequalities and Applications*, 2:269–273, 1998.

[Can83]   J. F. Canny. Finding edges in lines and images. Technical Report AI-TR-70, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, June 1983.

[CCN11]   V. Caselles, A. Chambolle, and M. Novaga. *Handbook of Mathematical Methods in Imaging*, chapter Total Variation in Imaging, pages 1016–1057. Springer, 2011.

[CFK04]   G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations Of Computational Mathematics*, 5:1–58, 2004.

[Cha04]   A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1–2):89–97, March 2004.

[Cia88]   P. G. Ciarlet. *Mathematical Elasticity. Volume I: Three-Dimensional Elasticity*. North-Holland, Amsterdam, 1988.

[CL97]   A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, (76):167–188, 1997.

[CS05]   T. F. Chan and J. Shen. *Image Processing and Analysis - Variational, PDE, wavelet, and stochastic methods*. SIAM, Philadelphia, 2005.

[Deu04]   P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*. Series Computational Mathematics. Springer, 2004.

[DG79]   E. De Giorgi. Convergence problems for functionals and operators. In *Proceedings of the International Meeting on Recent methods in non-linear analysis*, pages 131–188, Rome 1978, 1979.

[DH02]   P. Deuflhard and A. Hohmann. *Numerische Mathematik 1: Eine algorithmisch orientierte Einführung*. De-Gruyter-Lehrbuch. Walter de Gruyter, Berlin, 3rd edition, 2002.

[DR06]   M. Droske and W. Ring. A Mumford–Shah level-set approach for geometric image registration. *SIAM Journal of Applied Mathematics*, 66(6):2127–2148, 2006.

[ET99]   I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*, volume 28 of *Classics in Applied Mathematics*. SIAM, Philadelphia, PA, english edition, 1999. (Translated from French).

[Eva08]   L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. AMS, 2008.

[FHM00]   J.M. Fitzpatrick, D.L.G. Hill, and C.R. Maurer. *Handbook of Medical Imaging*, volume 2. SPIE Press, July 2000.

[FJSY09] M. Fuchs, B. Jüttler, O. Scherzer, and H. Yang. Shape metrics based on elastic deformations. *Journal of Mathematical Imaging and Vision*, 35(1):86–102, 2009.

[FKPP10] S. Fürtinger, S. L. Keeling, G. Plank, and A. J. Prassl. Registration of edge sets for mapping a Purkinje fiber network onto an endocardium. Technical report, Institute for Mathematics and Scientific Computing, Karl–Franzens University, Graz, Austria, 2010.

[FKPP11] S. Fürtinger, S. L. Keeling, G. Plank, and A. J. Prassl. Elastic registration of edge sets by means of diffuse surfaces - with an application to embedding Purkinje fiber networks. In Leonid Mestetskiy and José Braz, editors, *VISAPP*, pages 12–21. SciTePress, 2011.

[FKPP12] S. Fürtinger, S. L. Keeling, G. Plank, and A. J. Prassl. *Deformation Models: Tracking, Animation and Applications*, volume 7 of *Lecture Notes in Computational Vision and Biomechanics*, chapter Elastic Registration of Edges Using Diffuse Surfaces. Springer, 2012.

[GLV09]   J B Garnett, T M Le, and L A Vese. Some variational problems from image processing. *Image Processing*, pages 1–22, 2009.

[GR80]   V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations*. Springer, Berlin, 1980.

[GR05]   C. Großmann and H.-G. Roos. *Numerische Behandlung partieller Differentialgleichungen*. Teubner, Wiesbaden, 3rd edition, 2005.

[GX02]   J. Gallier and D. Xu. Computing exponentials of skew-symmetric matrices and logarithms of orthogonal matrices. *International Journal of Robotics and Automation*, 17(4):1–11, 2002.

[Hal74]   P. Halmos. *Measure Theory*. Springer, 1974.

[HB06]   W. Hill and R. A. Baldock. The constrained distance transform: Interactive atlas registration with large deformations through constrained distances. In *DEFORM'06 - Workshop on Image Registration in Deformable Environments*, 2006.

[HKR93]   D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, Sep. 1993.

[HKR$^+$10] M. Hofer, S. L. Keeling, G. Reishofer, M. Riccabona, M. Aschauer, and R. Stollberger. DCE-MRI nonrigid kidney registration. In *Proceedings of the ISMRM - 18th Scientific Meeting & Exhibition*, Stockholm, Sweden, May 2010. ISMRM.

[Jai89]   A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall Information and Systems Sciences Series. Prentice Hall, Englewood Cliffs, NJ, 1989.

[JKF01]   O. Jesorsky, K. Kirchberg, and R. Frischholz. Robust face detection using the Hausdorff distance. In *Third International Conference on Audio- and Video-based Biometric Person Authentication*, Lecture Notes in Computer Science, LNCS-2091, pages 90–95, Halmstad, Sweden, June 6–8 2001. Springer.

[JOP$^+$]   E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python. `http://www.scipy.org/`, 2001–.

[Kee10]   S. L. Keeling. Registration of modulated image sequences based upon locally rescaled connected segments. Technical report, Institute for Mathematics and Scientific Computing, University of Graz, Graz, Austria, 2010.

[Kel99]   C. T. Kelley. *Iterative Methods for Optimization*. Number 18 in Frontiers in Applied Mathematics. SIAM, Philadelphia, 1999.

[KKP12]   T. D. Katsaounis, S. L. Keeling, and M. Plexousakis. Consistent discretizations for vanishing regularization solutions to image processing problems. *Submitted*, 2012.

[KKS09]   C. Knauer, K. Kriegel, and F. Stehn. Minimizing the weighted directed Hausdorff distance between colored point sets under translations and rigid motions. In *FAW '09: Proceedings of the 3d International Workshop on Frontiers in Algorithmics*, pages 108–119, Berlin, Heidelberg, 2009. Springer-Verlag.

[KR05]   S.L. Keeling and W. Ring. Medical image registration and interpolation by optical flow with maximal rigidity. *Journal of Mathematical Imaging and Vision*, 23(1):47–65, 2005.

[Kra01]   S.G. Krantz. *Function Theory of Several Complex Variables*. AMS Chelsea Publishing Series. American Mathematical Society, 2001.

[KS02]   S. L. Keeling and R. Stollberger. Nonlinear anisotropic diffusion filters for multiscale edge enhancement. *Inverse Problems*, 18:175–190, Jan. 2002.

[Lue69]   D.G. Luenberger. *Optimization by Vector Space Methods*. Series in Decision and Control. John Wiley & Sons, 1969.

[MAT10]   MATLAB. *version 7.11.0.584 (R2010b)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[Mes06]   D. Meschede. *Gerthsen Physik*. Springer, 23rd edition, 2006.

[MFTM01]   D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8th International Conference on Computer Vision*, volume 2, pages 416–423, July 2001.

[Mod04]   J. Modersitzki. *Numerical Methods for Image Registration*. Oxford Science Publications, 2004.

[MS89]   D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[NW00]   J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Series in Operation Research and Financial Engineering. Springer, 2nd edition, August 2000.

[PRR02]   N. Paragios, M. Rousson, and V. Ramesh. Matching distance functions: A shape-to-area variational approach for global-to-local registration. In *Proceedings of the 7th European Conference on Computer Vision-Part II*, 2002.

[PSRS99]   W. Peckar, C. Schnörr, K. Rohr, and H. S. Stiehl. Parameter-free elastic deformation approach for 2d and 3d registration using prescribed displacements. *Journal of Mathematical Imaging and Vision*, 10(2):143–162, 1999.

[RE95]   P. Rosin and T. Ellis. Image difference threshold strategies and shadow detection. In *Proceedings of the British Machine Vision Conference*, pages 347–356. BMVA Press, 1995.

[ROF92]   L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Journal of Physics D: Applied Physics*, 60:259–268, 1992.

[RS75]   M. Reed and B. Simon. *Methods of Modern Mathematical Physics*, volume 2: Fourier Analysis, Self Adjointness. Academic Press, 1975.

[Sch30]   J. Schauder. Der Fixpunktsatz in Funktionalräumen. *Studia Mathematica*, (2):171–180, 1930.

[Sch81]   L. Schumaker. *Spline Functions: Basic Theory*. Wiley, New York, 1981.

[She05]   J. Shen. Gamma-convergence approximation to piecewise constant Mumford–Shah segmentation. In Jacques Blanc-Talon, Wilfried Philips, Dan C. Popescu, and Paul Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems - ACIVS*, volume 3708 of *Lecture Notes in Computer Science*, pages 499–506. Springer, 2005.

[Sho94]   R. E. Showalter. *Hilbert Space Methods for Partial Differential Equations*. Electronic Monographs in Differential Equations. Electronic reprint of the 1977 original, San Marcos, Texas, 1994.

[SL74]   L.A. Shepp and B.F. Logan. The Fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, June 1974.

[TA77]   A.N. Tikhonov and V.Y. Arsenin. *Solutions of Ill-Posed Problems*. Winston and Sons, Washington D.C., 1977.

[vRD12]   G. van Rossum and F. L. Drake, editors. *Python Reference Manual*. Python Software Foundation, `http://docs.python.org/ref/ref.html`, 2012.

[Zaj83]   L. Zajíček. Differentiability of the distance function and points of multi-valuedness of the metric projection in Banach space. *Czechoslovak Mathematical Journal*, 33(2):292–308, 1983.