Westfälische
Wilhelms-Universität
Münster

Variational Methods for Combined Image and Motion Estimation

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich
Mathematik und Informatik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

Fach: Mathematik

vorgelegt von:
Sebastian Suhr
aus Lüneburg
- Mai 2015 -

Dekan:	Prof. Dr. Martin Stein
Erster Gutachter:	Prof. Dr. Martin Burger
Zweiter Gutachter:	Prof. Dr. Jan Modersitzki
Tag der mündlichen Prüfung:	22.6.2015
Tag der Promotion	

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit zwei Variationsmodellen für simultane Bild- und Bewegungsschätzung. Nach Einordnen dieser Probleme in den Kontext medizinischer schlecht gestellter inverser Probleme und Vorstellen genereller Konzepte für die (numerische) Lösung, wenden wir uns zwei konkreten Problemen zu.

Wir leiten einen verallgemeinertes Funktional für Registrierung mit Rausch-Modellierung her, indem wir ein zugrunde liegendes echtes Bild als zusätzliche Variable in das Modell mit aufnehmen. Indem wir verschiedene Tranformationsoperatoren als lineare Operatoren auffassen können wir für fixe Transformationen ein optimales Bild berechnen und dieses in das Funktional einsetzen. Dies führt zu speziell auf Rausch- und Transformationsmodell zugeschnittene Distanzmaße im Registrierungsterm. Wir weisen nach, dass für den Masse-erhaltenden Transformationsoperator Existenz für eine breite Klasse von Rauschdistanzen eines Minimierers gezeigt werden kann. Im Intensitäts-konstanten Fall sind die entstehenden Distanzmaße jedoch konkav in der Funktionaldeterminante der Transformation, was den Existenznachweis mit Hilfe der direkten Methode der Variationsrechnung verhindert. Zum Abschluss vergleichen wir das Modell auf synthethischen Daten mit bestehenden Registrierungsalgorithmen.

Als zweites leiten wir aus Bayesianischer Modellierung ein Funktional zur bewegungskorrigierten Rekonstruktion her. Für dieses Funktional zeigen wir Existenz eines Minimierers unter der Voraussetzung, dass die Transformation, die das zu rekonstruierende Bild verformt, injektiv ist. Wir diskutieren dabei, wie diese Voraussetzung erfüllt werden kann. Nach einem Vergleich mit Standard Rekonstruktions- und Bewegungskorrekturalgorithmen auf einem künstlichen Entzerrungsbeispiel testen wir die vorgestellte Rekonstruktionsmethode noch für Soft- und Hardware-Phantom Daten.

Abstract

This thesis is subsequently concerned with two variational models for simultaneous image and motion estimation. After classifying these problems in the context of ill-posed problems in medical imaging, we focus on the presentation of (numerical) solution methods for the latter.

We start by deriving an unified framework for image registration with noise modelling by incorporating the underlying true image as additional unknown. By interpreting the motion as a linear operator acting on the image we are able compute an optimal image for any fixed transformation, which we plug into the functional. In doing so we obtain specific distance measures for any choice of transformation operator and data fidelity corresponding to the noise pertubation. We give an existence result for the mass-preserving transformation operator for a wide range of data fidelities, while the functional for the intensity-constant case turns out to be concave with respect to the Jacobian determinant of the transformation gradient, which prevents verifying the existence of a minimizer with the direct method in the calculus of variations. We conclude by comparing the model with other registration methods on artificial data.

The second model results from Bayesian Modelling for motion-corrected reconstruction. We show existence of a minimizer for the resulting functional for MAP estimation, provided that the transformation, which deformes the image to be reconstructed, is injective. We discuss two ways to incorporate injectivity as a constraint in the model. After comparing the proposed method with reconstruction and motion estimation algorithms on an artificial deblurring example, we test the proposed method on soft- and hardware phantom data.

Acknowledgments

I would like to thank

- Martin Burger and Jan Modersitzki for providing the opportunity to work on this interesting and challenging subject and always giving excellent guidance and support.
- Friedrich Suhr for carefully proof reading this thesis multiple times.
- Eva-Maria Brinkmann for proof reading and providing many synonyms for my favorite English phrases.
- Dirk Mannweiler for proof reading, data courtesy and support with the EMRecon toolbox countless times.
- Daniel Tenbrinck for proof reading and showing me countless tricks how to produce results and visualizations efficiently.
- Steffen Volkmann for proof reading and counting the brackets in my computations to ensure bracket-consistency.
- Christoph Brune, Lars Ruthotto, Benedikt Wirth and Frank Wübbeling for providing valuable advice.
- Hendrik Dirks, Lena Frerking, Constantin Heck, Thomas Polzin, Michael Schaefer and Matthias Schlottbom for inspiring discussions.
- All colleagues from the Institute of Mathematics and Image Computing, Lübeck as well as Institute for Computational and Applied Mathematics Münster for providing excellent working atmosphere and solving countless small and big problems.
- My parents Cornelia and Friedrich, my brother Fabian and the rest of my family and friends for always supporting me.
- Last but not least Rebecca Grothe for always believing in me, even when I did not.

Additionally i acknowledge the support of the Deutsche Forschungsgemeinschaft via the Project MO-1053/2-1.

Contents

1 Motivation

In this chapter we give a short motivation for the combined reconstruction and motion estimation methods we present throughout this thesis. We start by describing the practical relevance of reconstruction problems and draw the link to motion estimation.

The usage of X-radiation in medical imaging can be traced back to the late 19th century, when it was discovered by Wilhelm Conrad Röntgen.

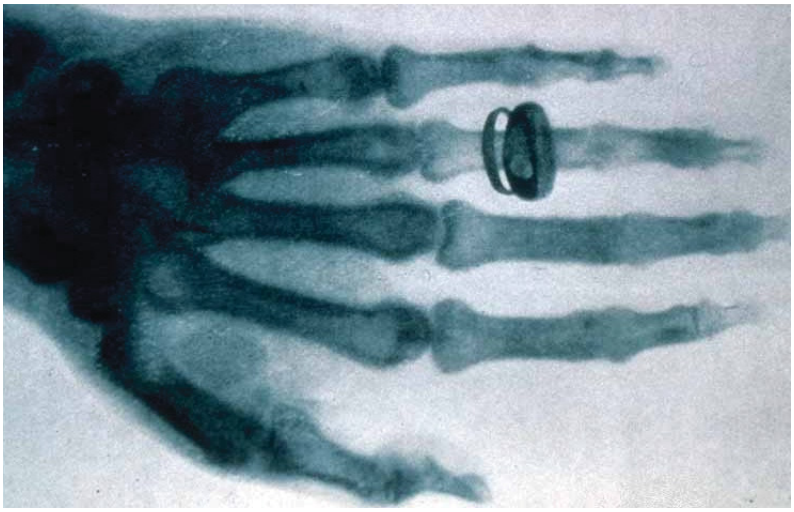


Figure 1.1: Early X-ray scan performed by Wilhelm Conrad Röntgen in 1896 [126].

Being initially a method for generating two dimensional images, it took around 70 years till 3D.imaging with X-radiation was possible via computerized tomography. Ironically Cormack [38] and Hounsfield [78], the two pioneers in computerized tomography, were not aware of the fact that the problem of determining a function by its line integrals was already solved by Johann Radon in 1917 [118]. Since computerized tomography only yields structural information (tissue and bones), there is also need for imaging techniques, such as for example emission tomography, yielding metabolic information.

The key idea in these emission tomography methods is to inject a radioactive tracer into the region to be studied to deduce abnormalities in the accumulation of the tracer via the emitted radiation. Since tumor cells often have an increased need for glucose, these cells can often be seen in emission tomography imaging [150]. Because a standard emission tomography scan might take several minutes the recorded data are often corrupted by external as well as internal patient movement. A common approach to alleviate these difficulties is to classify the data in several different stages of motion, the so called gates [28].

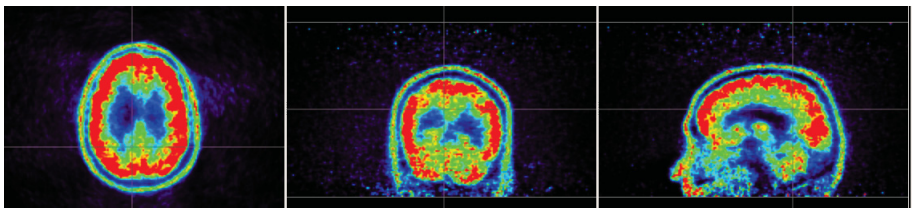


Figure 1.2: Gated reconstruction of PET head data. The noise corruption is clearly visible. Data courtesy of Dirk Mannweiler (EIMI Münster).

Naturally only a fraction of the complete data is available for each gate, which results in images with a low signal to noise ratio. To increase this ratio one could simply inject a greater tracer dose, but this would result in a higher radiation exposure for the patient as well as higher costs for the tracer. Thus one often tries to improve the image quality by motion-correction techniques [63]. We roughly divide reconstruction with motion-correction into three different classes:

- Gating and single gate reconstruction, e.g. [28].
- Aligning and averaging several reconstructed images, e.g. [63].
- Incorporating motion correction into the reconstruction process, e.g. [96].

Before we compare these three classes, we shortly outline some basic assumptions and reconstruction techniques in PET-imaging.

Usually the measured data in PET-imaging is assumed to be the Radon transform corrupted by Poisson noise [40]. For solving problems of the form

$$Ku = f \tag{1.1}$$

Shepp and Vardi proposed the expectation maximization (EM) algorithm to solve problem (1.1) with f being corrupted by Poisson noise [137]:

1. Set $u_0 \geq 0$.
2. Set $u_{k+1} = \frac{u_k}{K^*1} K^* \left(\frac{f}{Ku_k} \right)$.
3. Repeat step 2 until the maximum iteration number is reached.

Unfortunately in the case of noisy data, said noise is amplified in the iteration process and thus the EM-algorithm needs to be stopped in time [132]. Thus regularized versions of the EM algorithms have been developed; see e.g. [132, 131, 22] for expectation maximization with total variation (TV) regularization. Image deblurring, which is a task in medical imaging as well as microscopy, can be modeled in a similar way [89, 29]. We illustrate the potential of combined reconstruction and motion-correction with an artificial deblurring problem:

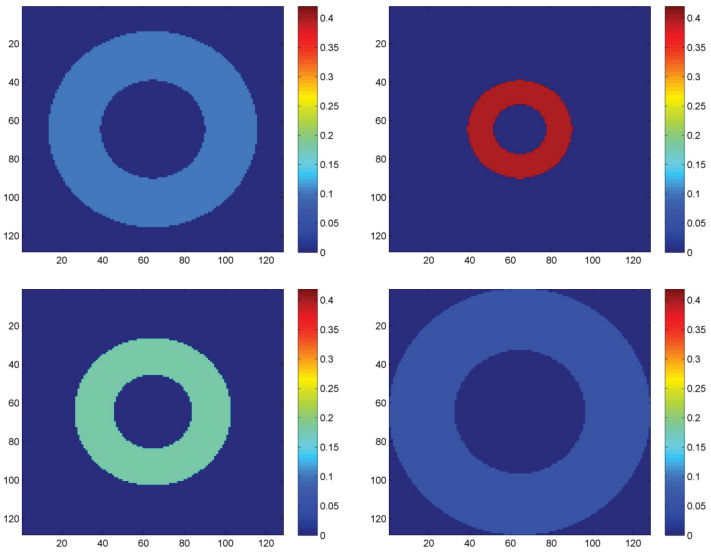


Figure 1.3: Ground truth data in different stages of compression as an artificial reconstruction problem; the reference gate to be reconstructed is on the top left. The same amount of activity is stored in each image. Exact aligning can be done via (mass-preserving) affine-2D transformation.

Since the total amount of activity is the same in each image, our reference gate can be matched exactly to each other image with a (mass-preserving) affine linear transformation. We apply the MATLAB disc-filter to obtain the blurred data :

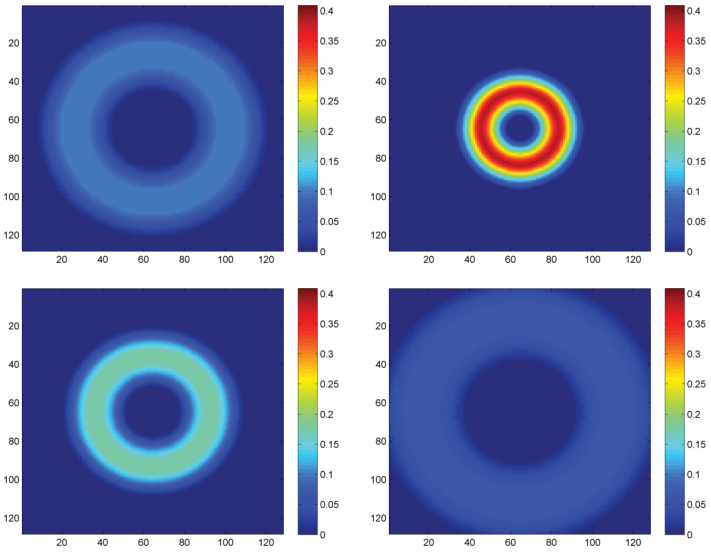


Figure 1.4: Blurred versions of the images in Figure 1.9 are generated with the MATLAB disc filter. Note that the images lose their sharp edges.

To simulate the corruption by Poisson noise occurring, e.g., in emission tomography and fluorescence microscopy, we add Poisson noise:

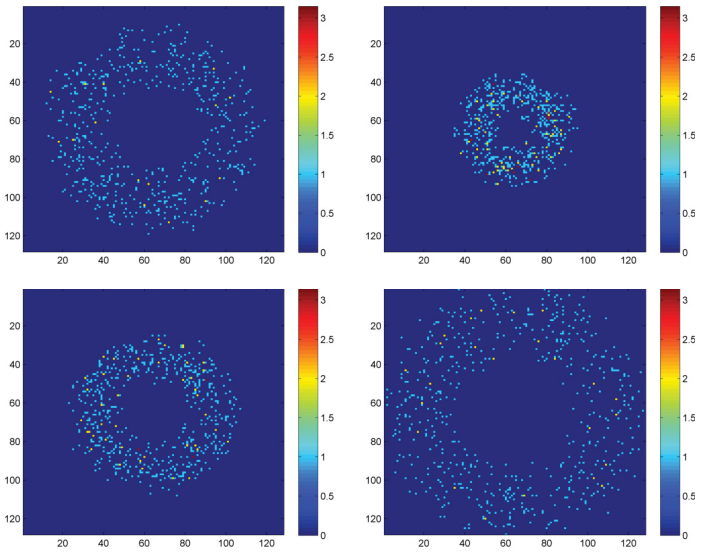


Figure 1.5: The blurred images in Figure 1.4 corrupted with Poisson noise. Note that the Poisson distribution only takes non-negative integer values and thus there is massive loss of information.

We start by comparing the three classes of reconstruction methods now by using the classical EM algorithm. By using the motion-vectors, which align the ground truth images exactly, we expect the motion-corrected method to yield better reconstruction than the averaging of reconstructed images, since the motion-corrected reconstruction grants one reconstruction from full data instead of averaging four incomplete reconstructions:

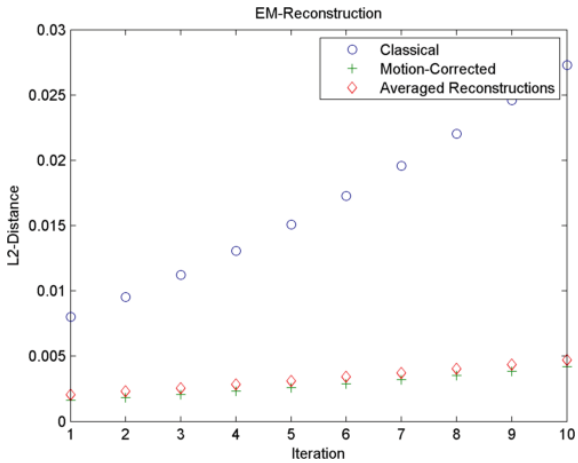
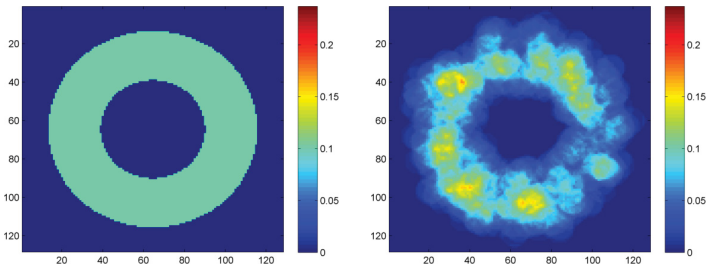
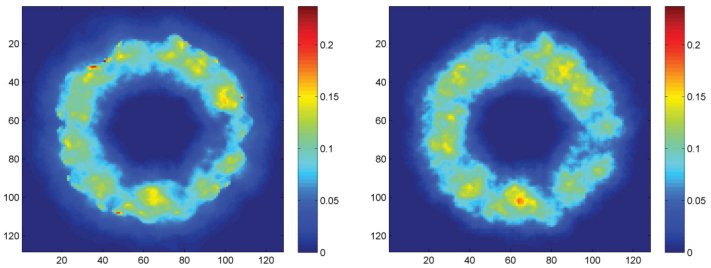


Figure 1.6: Reconstruction error versus the iterations of the classical EM algorithm. Note that the noise amplification is lower for the two motion-corrected methods.

The reconstruction errors are as we expected, but nevertheless we show the best reconstructions for each method:



(a) Ground Truth (b) EM



(c) av. EM (d) m.c. EM

Figure 1.7: Ground truth image and EM reconstructions for the three methods. The improvement by incorporating motion information is clearly visible.

Since the edges get blurred in all of the reconstructions we employ TV-regularization:

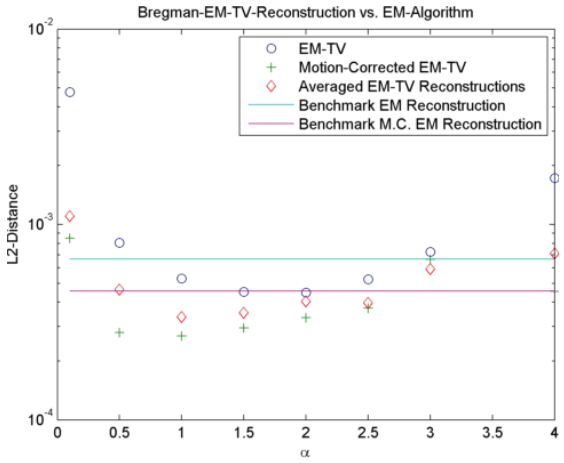
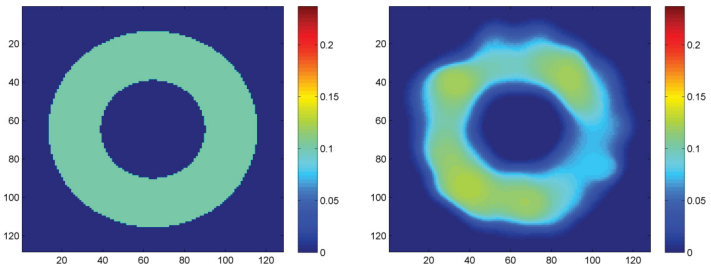
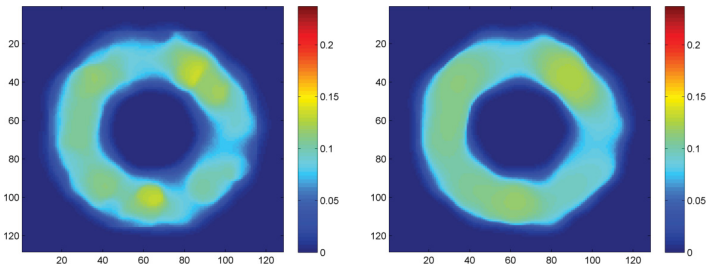


Figure 1.8: Reconstruction error versus the parameter of the Bregman EM-TV reconstruction with classical (motion-corrected) EM reconstructions as benchmark. Again the motion-corrected reconstruction performs best.

Before we start discussing this comparison, we give visual impression of the reconstruction results:



(a) Ground Truth (b) EM-TV



(c) av. EM-TV (d) m.c. EM-TV

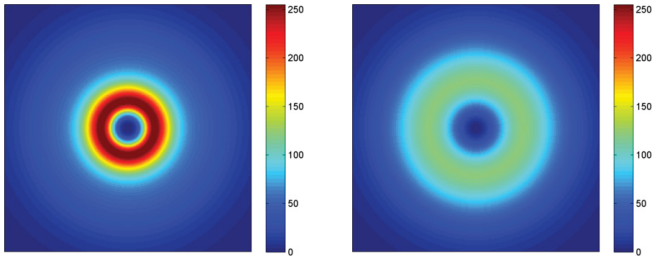
Figure 1.9: Ground truth image and EM-TV reconstructions for the three methods. The regularization grants a much more homogeneous ring for all methods. The motion-corrected reconstruction almost completely removes the fading effect on the right side of the ring.

We saw that incorporating motion information improved the reconstruction result for the unregularized and regularized EM algorithm. Furthermore incorporating the motion information directly into the reconstruction process was superior to averaging reconstructed images, since the motion-corrected reconstruction improves the SNR of the reconstructed image, while by averaging only images with poor signal to noise ratio are averaged. The motion-corrected EM reconstruction was even hardly bested by a TV-regularized single Gate reconstruction (compare Figure 1.8).

The artificial reconstruction problem we outlined above showed us, that incorporating the motion-information directly into the reconstruction process is superior to the other two classes. Note that we performed the motion-correction with known motion-vectors, which guaranteed exact matching of the ground truth images. In practical applications these motion-vectors are generally not available, thus motion estimation has to be performed.

Besides from enhancing the quality of reconstructed images, motion estimation is used for several other medical tasks, as we illustrate with the following example. In order to control the size of tumors in cancer treatment several follow up scans are performed. Naturally these images are not exactly aligned due to patient movement or other influences. Thus motion correction is often performed to quantify the tumor residue [138]. There are several other tasks in medical imaging, where motion-correction via image registration is performed, see for example [95] or more recently [112] for a survey on this topic.

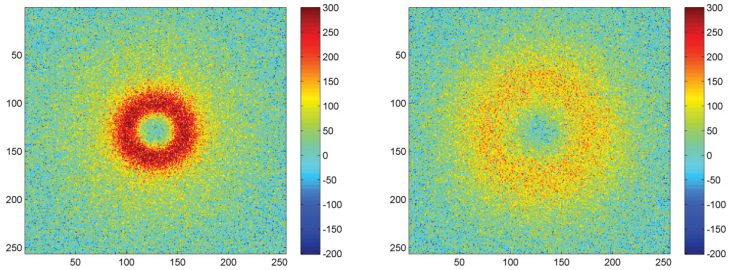
Unfortunately, the digital images acquired by medical imaging systems are usually corrupted by noise [23]. Naturally, this noise corruption is challenging for registration of these images. We will illustrate this fact with an artificial example, which we study in more detail in Section 6.5. Consider the following (mass-preserving) registration problem:



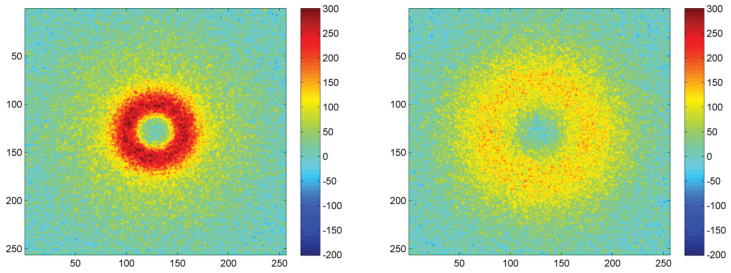
(a) Reference (b) Template

Figure 1.10: Ground truth data for mass preserving registration: Intensity distributed over a wide area (template) and the same amount concentrated in a small circle (reference).

To simulate noise corruption of this images, we add Gaussian noise with mean zero and standard deviation $\sigma = 50$. A common approach to deal with this noise in image processing is filtering the images with a Gaussian kernel:



(a) Noisy reference (b) Noisy template



(c) Filtered reference (d) Filtered template

Figure 1.11: Noisy version of the registration problem in Figure 1.10 (top row). The Gaussian filtered versions can be found in the bottom row.

We compare the quality of the registration on the three data sets now by transforming

the ground truth with the found transformation for various values of the regularization parameter α (See Section 6.5 for more details on the evaluation).

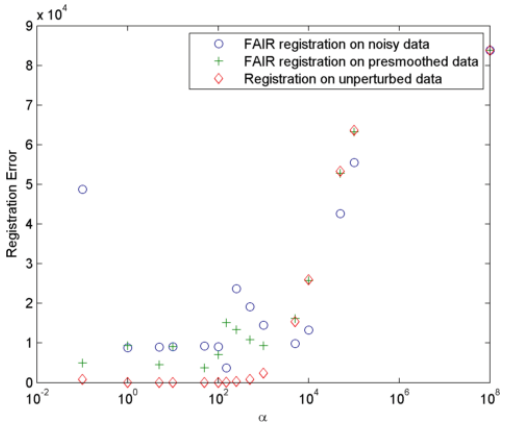


Figure 1.12: Registration for motion estimation performed on noisy, presmoothed and unperturbed data. The ground registration on unperturbed data provides a good solution for a wide parameter range, while the presmoothed registration performs slightly better than the registration on noisy data.

As expected we see that performing motion estimation on noise free data is much more stable then registering noisy images. Even filtering only leads to a small improvement of the estimated transformation. This observation raises the question, whether incorporating information on the noise characteristics can improve the quality of the image estimation. To our best knowledge this subject has not yet been investigated in a general framework, but there are some studies for specific noise and transformation models. See e.g. [3] for a statistical framework in the case of Gaussian noise. Thus in Chapter 6 we aim to derive a framework, which is independent of the noise corruption as well as of the transformation model.

Going back to the case of estimating motion in order to enhance the quality in image reconstruction, we expect that the quality of the reconstructed image and estimated motion field are linked. Thus we hope that we can improve the quality of an reconstructed image by performing motion-corrected reconstruction; this improved image can then be used to improve the motion estimation. Motivated by this expectation we formulate a coupled model for image reconstruction and motion estimation in Chapter 7.

In conclusion we can say, that incorporating exact motion information improved reconstruction from noisy data, as well as performing registration on noise free data led to a much more stable registration result. While this conclusion is not surprising at all, we point it out, explicitly to motivate the combined motion and image estimation models we will present throughout this thesis. Note that both problems are similar in the modeling, but can be viewed from different angles: In Chapter 6 we use an implicit image estimation to incorporate noise information into a registration framework, while we use simultaneously estimated motion in order to enhance reconstruction quality in Chapter 7.

2 Structure of the Thesis

We start by presenting some basic definitions and results in the first part of Chapter 3. Since we are dealing with the concept of mass-preservation the transformation theorem is a fundamental tool in the analysis we carry out later in this thesis. In order to generalize the classical version (Theorem 3.1.13) we present the extension of classical infinitesimal calculus to functions defined on spaces equipped with integral norms in Section 3.2. The key idea is to assign point values for functions via averaging; as it turns out those values are well-defined almost everywhere. These results allow us to generalize the transformation theorem to functions, which are not necessarily injective and differentiable (Theorem 3.2.7). Following these ideas of approximate differentiability we present the generalization of diffeomorphisms to a certain class of Sobolev functions. Although the notion of weak diffeomorphisms originated in the theory of Cartesian currents, we are able to restrict ourselves to subsets of Sobolev mappings, which are also weakly diffeomorphic.

The key tool in the analysis of our variational problems is the direct method in the calculus of variations (Section 3.4): After stating some basic definitions for the analysis of functionals we are able to give sufficient conditions for the existence of a minimizer: Compactness of the sublevel set ensures that any infimizing sequence has a convergent subsequence, while lower semicontinuity guarantees that the limit of this subsequence minimizes the functional. While the compactness of sublevel sets often can be deduced by coercivity properties and the theorem of Banach-Alaoglu (Theorem 3.4.7), verification of lower semicontinuity needs more caretaking (compare the analysis sections in Chapters 6 and 7).

The last part of this basic theoretical chapter is devoted to introducing linear operators and stating some basic properties. While linear operators are useful in the modeling of transformations, projection operators in tomography are often assumed to be compact. Therefore we conclude the survey on linear operators by stating some properties of compact operators, which transform bounded into totally bounded sets.

Note that the tools presented in Sections 3.1, 3.4 and 3.5 are often used in the analysis of image processing problems and thus these Sections can be skipped by the experienced reader. To our best knowledge there are no other works using results from Sections 3.2 and 3.3 in the context of image processing and thus we recommend those sections to all readers.

Having collected and illustrated these basic definitions and tools we focus on inverse problems in imaging in Chapter 4. We start by outlining the basic concept of forward and inverse problem, as well as ill-posedness of said problems. Afterwards we present a

short survey on regularization and statistical methods to deal with ill-posed problems.

In the second part of this chapter we put reconstruction and motion estimation problems in the context presented before. After giving short examples of reconstruction problems occurring in medical imaging we present the basic definition and properties of the widely used TV-regularization.

In the section on image registration, we state the standard form of the image registration problem. Following Modersitzki's definition [100] this problem is formed by several building blocks. We shortly highlight some examples for transformation operators, distance measures and regularizers and illustrate the difference between parametric and non-parametric image registration.

This chapter serves only as collection of some basic concepts used to deal with ill-posed problems, reconstruct images and estimate motion via registration. Thus this chapter can be skipped by the experienced reader, but we point out, that the difference between mass-preserving and intensity constant transformation operator for image registration has severe impact in Chapter 6 and should be understood by every reader interested in the framework for registration with noise modeling.

Numerical methods to deal with the (variational) problems arising from Chapter 4 are presented in Chapter 5. Usually one distinguishes between two types of optimization algorithms: Discretize-then-Optimize leads to a discretized version of the functional to be minimized, while in Optimize-then-Discretize methods analytical optimality conditions for said functional are solved numerically. Since we use Discretize-then-Optimize methods for both models we present later, the focus in this chapter is on these methods, nevertheless some basic ideas for Optimize-then-Discretize algorithms are outlined in Section 5.2.

After presenting ideas on how to discretize the building blocks of registration and denoising functionals, we concentrate on how to optimize these discretized versions. While in trust region methods, the function is approximated by a (usually quadratic) model function and then optimized in a trust region, where this model fit is sufficient, in line search methods the function is optimized along a specific search direction. We describe some methods how to determine this search direction, including the modified BFGS update we use.

This survey is concluded by a short overview on multilevel methods: Originated from multigrid solvers for partial differential equations, the idea is to minimize the function for a coarse discretization in order to obtain a starting value for the finer level. In doing so one hopes to reduce the probability of obtaining unwanted local minima. We complete this chapter by shortly outlining how this multilevel approach is affected by data corrupted with Gaussian noise.

Again this chapter serves as a summary of several well known definitions and facts and can therefore be skipped by readers experienced in this subject. However the modified BFGS update (5.18), which grants convergence for a wide range of non-convex objective

functions, and Example 5.1.1, which illustrates how the variance of noisy data is changed in a multilevel framework should be well-understood for the next chapters.

Developing a unified framework for motion estimation in noisy image sequence is the topic of Chapter 6. We start by deriving a maximum a posteriori estimator for a Bayesian modeling of determining true images and motion from noisy images recorded for a certain scene. By restricting this presentation to the intensity-constant and mass-preserving transformation operator we deduce a framework for image registration incorporating noise information.

By modeling the motion between two noisy images we obtain a functional in two unknowns: The true image corresponding to one noisy image and the motion between those images. By employing now regularization on this true image, we can directly optimize the functional for each fixed motion and plug this solution map into the functional to obtain a registration functional with implicit image estimation.

As the actual computation of the solution map and obtaining a functional depending only on the forward transformation consists of some technical computations, we only present these for the case of Gaussian noise in Section 6.2. For the sake of completeness these computations can be found in the Appendix 9.1 for Poisson and speckle noise.

We aim to prove the existence of a minimizer for these functionals via the direct method in the calculus of variations. While the compactness of sublevel sets is a direct consequence of employing suitable regularization on the motion, we need to verify polyconvexity, i.e. convexity with respect to the Jacobian determinant of the tranformation, in order to guarantee lower semicontinuity of the registration functional. Interestingly we are able to prove that functionals resulting from the mass-preserving transformation operator are convex with respect to the Jacobian determinant (Theorem 6.3.2), while functionals arising from intensity-constant transformation modeling are concave with respect to the Jacobian determinant (Theorem 6.3.4). Thus existence of a minimizer for the mass-preserving transformation operator is can be shown (Theorem 6.3.3), while we are not able to establish a similar result for the intensity-constant transformation operator.

We outline the numerical framework we use to minimize the resulting discretized objective functions. The focus in this presentation lies on deriving a modified objective function and optimization method, in order to use as much routines from the FAIR toolbox [101] as possible. Followed by this the registration method is evaluated thoroughly on an artificial example, while application towards real data is given by registering PET images from the XCAT software phantom.

The open questions discussed in the closure of this chapter can be divided in three blocks: Injectivity, convergence properties and noise modeling. Since we assumed our transformations to be injective for computing the adjoint transformation operators, we outline a registration framework with (injective) boundary conditions. Additionally we sketch the consequences in the modeling, if we allow the transformations to be non-injective. After computing the adjoint operators in this case, we derive a distance

measure for Gaussian noise with the mass-preserving transformation operator as an example.

In Chapter 7 we derive and analyze a model for motion-corrected reconstruction. After developing a nonlinear variational approach to this problem by Bayesian modeling we deal with the analysis of the resulting functional. The backbone of this section is Theorem 7.2.1, which grants us weak L^1-convergence of a sequence of images after transformations. Besides the regularity properties of image and transformations, which are granted by the regularization we imposed, we also needed that the number of roots to a point in the range of the transformation is bounded in the L^∞-sense. In order to satisfy this condition we give existence results for constraints guaranteeing injectivity of the transformation.

After showing that minimizing the functional is a regularization method in the sense of Definition 4.1.2, we introduce a numerical framework for minimizing the reconstruction functional. Again we use the Discretize-then-Optimize method; we pursue an alternating minimization scheme to tackle this optimization problem, which is defined on a product space. While reconstruction with incorporating motion information only changes the projection operator we use, minimizing with respect to the transformation leads to a registration problem. In contrast to the classical registration problem we assess the similarity of the transformed template in the detector domain instead of the image domain. While the reconstruction step is implemented using routines from the EMRecon [84] toolbox, the images estimation step is again performed with the help of FAIR [101].

In Section 7.4 we test the proposed method on three different datasets. After giving thorough evaluation and highlighting the potential on an artificial deblurring problem, we give a proof of concept for the usability in clinical applications by reconstructing noisy data from the XCAT-software phantom as well as the hardware phantom "Wilhelm".

The outlook on further research is devoted in outlining a weaker formulation of the problem, followed by a discussion on how the functional can be modified to incorporate attenuation correction as well as a priori information. We conclude this section by discussing the convergence properties of the minimization algorithm we chose and shortly sketching out an alternative, which grants stronger convergence properties.

In Chapter 8 we do a brief wrap up of the central questions arising in the analysis of both problems we presented in Chapters 6 and 7. After pointing out that the motion-corrected reconstruction framework can be seen as a generalization of the unified registration framework we derived in Chapter 6 we conclude by sketching out a proof that the Banach indicatrix of local diffeomorphisms is bounded in L^∞.

3 Theoretical Basics

Before dealing with the reconstruction and motion estimation topics, which we will investigate in the remainder of this thesis, we aim to present some basic tools, useful in the analysis for the models to be developed. We start by giving a short overview on the function spaces and mathematical tools we will use throughout this thesis. After stating these basic definitions and results, we will present some more specialized tools we need for the analysis carried out later in this thesis. We focus on a general summary of existing literature, but illustrate this summary with examples.

3.1 Function Spaces and Diffeomorphisms

Since the focus of this thesis lies on imaging problems we start with the following definition from [100] of an image:

Definition 3.1.1 (Image)**.** Let $\Omega \subset \mathbb{R}^d$ be bounded. A function $\rho : \Omega \to \mathbb{R}$ is called a density image, iff

- $\text{supp}(\rho) \subset \Omega$
- $0 \leq \rho(x) < \infty \qquad \forall x \in \Omega$
- $\int\limits_{\Omega} \rho^k \mathrm{d}x < \infty \qquad \text{for } k \geq 1$

The value $\rho(x)$ is called the intensity of ρ at x.
We can extend this definition to sequences of images $\rho(x,t)$ by requiring, that $\rho(x,t)$ is an image for each t.
Furthermore the definition above can be relaxed by allowing negative intensities as well.

As we see in the definition above a mathematical image is defined with some integrability constraints. We will now present some of the function spaces, which are useful in the analysis of imaging problems. A basic start for many definitions are the Lebesgue spaces, thus we shortly give the basic definitions from [2]. Formally the Lebesgue spaces are defined for a certain measure and a σ-algebra of measurable sets. In order to keep this presentation short we will not elaborate on possible choices for σ-algebrae and associated measures, but state only that we will use the Borel algebra and the Lebesgue measure throughout this thesis; compare e.g. [13] for a general course on measure and integration theory.

Definition 3.1.2 (L^p-norm)**.** Let $\Omega \subset \mathbb{R}^d$ and $u : \Omega \to \mathbb{R}$ and $0 < p \leq \infty$. For $0 < p < \infty$ we denote the L^p-norm with respect to the Lebesgue measure λ by:

$$\|u\|_p := \left(\int_\Omega |u(x)|^p \mathrm{d}\lambda(x) \right)^{\frac{1}{p}} \tag{3.1}$$

and for $p = \infty$ the p-norm is given by

$$\|u\|_\infty := \operatorname*{ess\,sup}_{x \in \Omega} |u(x)| \tag{3.2}$$

Consequently we define the space L^p as

$$L^p(\Omega) := \{u : \Omega \to \mathbb{R} \,|\, \|u\|_p < \infty\} \tag{3.3}$$

Furthermore there are the Lebesgue spaces of locally integrable functions:

$$L^p_{loc}(\Omega) := \{u | \int_K |u(x)|^p \mathrm{d}x < \infty \qquad \forall K \subset \Omega \text{ compact}\} \tag{3.4}$$

Remarks and Extensions. The triangle inequality for $\| \cdot \|_p$ is a consequence of the Minkowski inequality, which makes the p-norm a seminorm. For gaining a full norm one usually identifies functions only differing on zero sets by setting:

$$0_{L^p} := \{f : \Omega \to \mathbb{R}| \quad f = 0 \quad \text{a.e.}\} \tag{3.5}$$

This has the consequence that we deal with equivalence classes of functions. See for example [20] for more details.

In the case $p = 2$, the L^2 norm can be deduced from the L^2 inner product

$$\langle f, g \rangle_{L^2\Omega} := \int_\Omega f(x) g(x) \mathrm{d}x,$$

which makes L^2 a so-called Hilbert space. We will give this definition shortly:

Definition 3.1.3 (Hilbert spaces)**.** Let $\mathcal{X}$ be a space equipped with an inner product $\langle \cdot, \cdot \rangle_\mathcal{X}$. Iff $\mathcal{X}$ is complete with respect to the norm

$$\|x\|_\mathcal{X} := \sqrt{\langle x, x \rangle}$$

we call $\mathcal{X}$ a Hilbert space.

Another powerful tool in the analysis of Lebesgue functions is Hölder's inequality, which allows us to determine a sufficient criterion, whether a product of functions can be controlled in a certain Lebesgue spaces:

Theorem 3.1.4 (Hölder's inequality)**.** *Let $p, q, r \geq 1$ and $\frac{1}{p}+\frac{1}{q} = \frac{1}{r}$. Suppose $f \in L^p(\Omega)$, $g \in L^q(\Omega)$. Then $fg \in L^r(\Omega)$ and*

$$\|fg\|_r \leq \|f\|_p \|g\|_q \tag{3.6}$$

Proof. See e.g. [119, Thm. 3.1 (c)]. □

With these basics spaces, we will now define Sobolev spaces. These spaces consist of functions, which are differentiable in a generalized sense. We restrict ourselves to summarizing the central results from [45, 2]. Inspired by partial integration we can now define a weak form of derivatives:

Definition 3.1.5 (Weak derivative)**.** Let $u, w \in L^1_{loc}(\Omega)$ and α be a multiindex. We say that w is the α-th weak derivative of u, written

$$D^\alpha u = w \tag{3.7}$$

provided

$$\int_\Omega u D^\alpha \phi \mathrm{d}x = (-1)^\alpha \int_\Omega w\phi \mathrm{d}x \tag{3.8}$$

for all test functions $\phi \in C_c^\infty(\Omega)$.

Remarks and Extensions. If it exists a weak partial derivative is uniquely defined up to a zero set [45].

We give the definition of Sobolev spaces from [45]:

Definition 3.1.6 (Sobolev spaces)**.** Let $1 \leq p \leq \infty$ and k be a non-negative integer. Then the Sobolev space

$$W^{k,p}(\Omega)$$

consists of all functions $u \in L^1_{loc}(\Omega)$ such that for each multiindex α with $|\alpha| \leq k$ the weak partial derivative $D^\alpha u$ exists and belongs to $L^p(\Omega)$

Remarks and Extensions. In the case of $p = 2$ one usually writes $H^k(\Omega) := W^{k,2}$, since $W^{k,2}$ is a Hilbert space. Sobolev spaces can be defined analogously for mappings $u : \Omega \to \mathbb{R}^{\tilde{d}}$ by requiring that each component function is in said space.

The Sobolev spaces are usually equipped with the following norm:

Definition 3.1.7 (Sobolev norm)**.** If $u \in W^{k,p}$ and $p < \infty$ we define its norm to be:

$$\|u\|_{W^{k,p}(\Omega)} := \left(\sum_{|\alpha| \leq k} \int_\Omega |D^\alpha u|^p \mathrm{d}x \right)^{\frac{1}{p}} \tag{3.9}$$

For the case of $p = \infty$ we set:

$$\|u\|_{W^{k,p}(\Omega)} := \max_{|\alpha| \leq k} \operatorname{ess\,sup}_{x \in \Omega} |D^\alpha u(x)| \tag{3.10}$$

Despite the fact, that $\partial\Omega$ is a $\mathbb{R}^d-$ zero set, boundary values can be determined for Sobolev functions [45].

Theorem 3.1.8 (Trace Theorem)**.** *Assume Ω is bounded and $\partial\Omega$ is C^1. Then there exists a bounded linear operator*

$$T : W^{1,p}(\Omega) \to L^p(\partial\Omega)$$

such that

$$Tu = u|_{\partial\Omega} \quad \textit{if} \quad u \in W^{1,p}(\Omega) \cap C^0(\Omega) \tag{3.11}$$

and

$$\|Tu\|_{L^p(\partial\Omega)} \leq C\|u\|_{W^{1,p}(\Omega)} \tag{3.12}$$

for each $u \in W^{1,p}(\Omega)$, with the constant C only depending on p and Ω.

The linear operator is usually called trace operator, while the assigned boundary values are known as Sobolev traces. Before giving criteria, when Sobolev functions belong to certain other spaces, we give the definition of the Hölder semicontinuity [45].

Definition 3.1.9. If $u : \Omega \to \mathbb{R}$ is bounded and continuous, we write

$$\|u\|_{C(\Omega)} := \sup_{x\in\Omega} |u(x)|.$$

Furthermore the Hölder seminorm of u is given by

$$[u]_{C^{0,\gamma}(\Omega)} := \sup_{x,y\in\Omega, x\neq y} \left\{ \frac{|u(x) - u(y)|}{|x - y|^\gamma} \right\}$$

Definition 3.1.10 (Spaces of Hölder-continuous functions)**.** The Hölder space

$$C^{k,\gamma}(\Omega)$$

consists of all functions $u \in C^k(\Omega)$ for which the norm

$$\|u\|_{C^{k,\gamma}(\Omega)} := \sum_{|\alpha|\leq k} \|D^\alpha u\|_{C(\Omega} + \sum_{|\alpha|=k} [D^\alpha u]_{C^{0,\gamma}(\Omega)} \tag{3.13}$$

is finite.

The following theorem from [45] provides useful information, whether a function $u \in W^{k,p}$ belongs to other spaces.

Theorem 3.1.11 (General Sobolev inequalities)**.** *Let Ω be a bounded open subset of $\mathbb{R}^d$ with C^1-boundary. Assume $u \in W^{k,p}(\Omega)$. If*

$$k < \frac{d}{p}$$

then

$$u \in L^q(\Omega) \qquad \text{with} \qquad \frac{1}{q} = \frac{1}{p} - \frac{k}{d}. \tag{3.14}$$

Moreover if

$$k > \frac{d}{p},$$

then there is u^* with $u^* = u$ a.e and

$$u^* \in C^{k-\left[\frac{d}{p}\right]-1,\gamma}, \tag{3.15}$$

where

$$\gamma = \begin{cases} \left[\frac{d}{p}\right] + 1 - \frac{d}{p} & \frac{d}{p} \notin \mathbb{N} \\ a \in (0,1) & \frac{d}{p} \in \mathbb{N} \end{cases}.$$

Proof. See [45, Chapter 5.6.4, Thm 6]. □

Having defined integral norms and spaces of integrable functions, we put our focus on a powerful tool for the analysis of the arising motion-correction problems. Since we are often concerned how the integral of functions after transformations changes, we start by defining a suitable class of such transformations:

Definition 3.1.12 (C^k-diffeomorphism)**.** Let $\Omega \subset \mathbb{R}^d$ be open and

$$y : \Omega \to \mathbb{R}^m \tag{3.16}$$

a mapping. Then y is called a C^k-diffeomorphism, if and only if y is bijective and y aswell as y^{-1} are of the class C^k.
If for each $x \in \Omega$ there exists a neighbourhood $U(x)$, such that y is a diffeomorphism between U and $y(U)$, then y is called a local diffeomorphism.

Remarks and Extensions. As a consequence of the implicit function theorem [104, Thm. 9.2] a C^1 function is a local diffeomorphism at $x \in \Omega$, provided

$$\det(Dy)(x) \neq 0. \tag{3.17}$$

In the case of $k = 1$ we simply call y a (local) diffeomorphism.

Having defined this class of suitable transformations, we can state the following tool from [104] to evaluate the integral of a transformed function.

Theorem 3.1.13 (Transformation Theorem for Diffeomorphisms)**.** *Let $\Omega \subset \mathbb{R}^d$ open and $y : \Omega \to y(\Omega)$ be a diffeomorphism. Then any function u is integrable, iff $u(y)\det(\nabla y)$ is integrable. Furthermore the following formula holds:*

$$\int_{y(\Omega)} u(z)\mathrm{d}z = \int_{\Omega} u(y(x))\det(\nabla y(x))\mathrm{d}x \tag{3.18}$$

Proof. See e.g. [104, Thm. 17.2]. □

As we will see in the next section, there is a version of the transformation theorem with weaker assumptions, which will become crucial, since our regularization functionals are not able to guarantee diffeomorphic motion estimates.

3.2 Infinitesimal Calculus for Integral Norms

In this section we generalize known definitions from the classical infinitesimal calculus to equivalence classes of functions in Lebesque- respectively Sobolev spaces (compare for example [57, 69, 60] for a further course on this matter). The final goal is to derive a version of the transformation theorem for integrals for non-diffeomorphic functions. Since we can not distinguish functions differing on zero sets, the classical definition of differentiability is not a feasible way, because we would like to obtain the same result for all representatives of the equivalence class. A natural way to define a coherent function value for all representatives is via averaging, which leads to the following definition.

Definition 3.2.1. Let Ω be a domain and $y \in L^1$. Then the set of points x for which $y_l(x)$ exists, such that

$$⨍_{B(x,r)} |y(z) - y_l(x)|\mathrm{d}z \to 0 \qquad \text{as} \qquad r \to 0 \tag{3.19}$$

is called the Lebesgue set $\mathscr{L}_y$, while the points in $\mathscr{L}_y$ are called Lebesgue points.

Remarks and Extensions.

- It is known that the complement of the Lebesgue points is a zero set [55, Thm. 2.19].
- For $y \in W^{1,1}_{loc}$ we can define the set of Lebesgue points $\mathscr{L}_{Dy}$ for the derivative Dy analogously.
- If y is a vector-valued function, then we say x is a Lebesgue point, iff it is a Lebesgue point for each component function.

We can generalize differentiability by the same idea [5]:

Definition 3.2.2 (Approximate differential for L^1_{loc})**.** Let $y \in L^1_{loc}(\Omega, \mathbb{R}^m)$ and let $x \in \mathscr{L}_y$; we say that y is approximately differentiable at x, iff there exists a $d \times m$ matrix L, such that

$$\fint_{B_r(x)} \frac{|y(z) - y_l(x) - L(z-x)|}{r} \mathrm{d}z \to 0 \qquad \text{as} \qquad r \to 0. \tag{3.20}$$

We present a generalized idea of the definition above. Although a standard course on approximate differentiability is given by [47], we rather use the definition by Whitney [147] in order to stick to the notation by Hajlasz [69], who formulated the generalized version of the transformation theorem for integrals we will present later. First we need to define a certain class of points.

Definition 3.2.3 (Density point)**.** Let $E \subset \mathbb{R}^d$ be a Lebesgue measurable set. We say that $x \in E$ is a density point of E with respect to the Lebesgue measure λ^d, iff

$$\lim_{r \to 0} \frac{\lambda^d(B_r(x) \cap E)}{\lambda^d(B_r(x))} = 1. \tag{3.21}$$

In other words this means, that part of the ball, which is not in E can be neglected. Now we can define approximate differentiability, by requiring that the points where the taylor expansion with help of the approximate differential is inaccurate is negligible in the sense of Definition 3.2.3.

Definition 3.2.4 (Approximate differential)**.** Let $y : \Omega \subset \mathbb{R}^m \to \mathbb{R}$ be a measurable function. $L = (L_1, ..., L_n)$ is said to be the approximate total differential of y at x_0, if for every $\varepsilon > 0$ the set

$$A_\varepsilon := \left\{ x \;\middle|\; \frac{|y(x) - y(x_0) - L(x - x_0)|}{|x - x_0|} < \varepsilon \right\} \tag{3.22}$$

has x_0 as a density point.

Remarks and Extensions.

- If the condition (3.22) holds, then x_0 is a density point of Ω and L is uniquely determined.
- The approximate differential is uniquely determined, if it exists. Furthermore each Sobolev function $y \in W^{1,p}_{loc}$ is approximately differentiable and the approximate differential equals the weak derivative a.e. [46, Chapter 6,Thm 4].
- Also this definition can be extended to vector-valued functions by applying it to each component.

With this at hand we can continue describing necessary definitions to derive a generalized version of the change of variables formula. A detailed description of generalized differentiability can be found e.g. in [147].
Since Sobolev functions can be changed on zero sets, we need to control the behavior of functions on zero sets, as the following example illustrates.

Example 3.2.1 (Redefining functions on zero sets)**.** For a short illustration of weird behaviour of functions on zero sets we recall the so called Cantor set $\mathscr{C}$: This set is constructed by starting with the interval $[0, 1]$ and iteratively wiping out the middle thirds; see Figure 3.1 for an illustration.

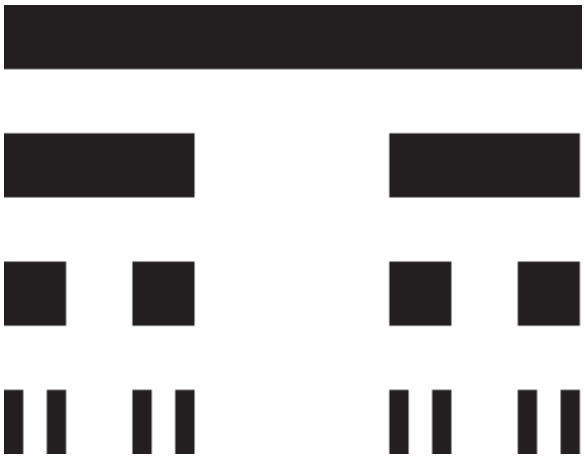


Figure 3.1: Visualization of the first four construction steps for the Cantor set.

The remaining points are all the points $x = \sum_i \frac{x_i}{3^i}$, $x_i \in \{0, 2\}$, which can be expressed without the digit 1 in the ternary expansion. It is well known that $\mathscr{C}$ is a zero set which can be mapped one-to-one to $[0, 1]$ with the so called Cantor function $f_{\mathscr{C}}$. The idea is to express $f_{\mathscr{C}}(x)$ by replacing the digit 2 with 1 and doing a binary instead of the ternary expansion [139].
We now take a look at two functions y_1 and y_2

$$y_1 : [0,1] \to \mathbb{R} \quad x \to x + 1$$
$$y_2 : [0,1] \to \mathbb{R} \quad x \to \begin{cases} x + 1 & x \in [0,1] \backslash \mathscr{C} \\ f_{\mathscr{C}} & x \in \mathscr{C} \end{cases}$$

Since the Cantor set is a zero set, it follows directly that $\|y_1 - y_2\|_p = 0$ for $1 \leq p < \infty$, but $\mathrm{Im}(y_1) = [1, 2] \neq [0, 2] = \mathrm{Im}(y_2)$. Thus y_2 blows the zero set $\mathscr{C}$ up to something substantial. Note that y_2 does only takes its Lebesgue value on $[0, 1] \setminus \mathscr{C}$.

To overcome difficulties arising from changing functions on a zero set, Nikolai Lusin imposed in his dissertation [94] the so called Lusin condition[1], also known as N-condition:

Definition 3.2.5 (Lusin's condition)**.** A mapping $y : \Omega \to \mathbb{R}^d$ satisfies Lusin's condition, iff for all:

$$E \subset \Omega,\ \lambda(E) = 0 \Rightarrow \lambda(y(E)) = 0 \tag{3.23}$$

[1]Historical remarks in this thesis are due to the encyclopedia of mathematics [74], if not explicitly said otherwise.

with λ denoting the Lesbesgue-measure.
This condition can be generalized to other measures as well.

Remarks and Extensions. One usually defines a Lusin representative $\tilde{y}$ by setting :

$$\tilde{y}(x) = \begin{cases} y_l(x) & x \in \mathscr{L}_y \\ z \quad \text{with} \quad z \notin \operatorname{Im}(y) & x \in \Omega \backslash \mathscr{L}_y \end{cases}. \tag{3.24}$$

In the following we always assume functions to be Lusin representatives, if not explicitly said otherwise.

Having surmounted these difficulties we need to take in account, that a non-diffeomorphic function may hit some points several times. For this problem Stefan Banach introduced the so called Banach indicatrix, which gives the number of roots to an equation [10]. This concept was generalized to discontinuous functions by Lozinski [93] and into higher dimensions by Kronrod [88] and Vitushkin in his master thesis [145]:

Definition 3.2.6 (Banach indicatrix)**.** Let $y : \mathbb{R}^d \to \mathbb{R}^m$, $E \subset \mathbb{R}^d$. The Banach indicatrix

$$N_y(\cdot, E) : \mathbb{R}^m \to \mathbb{N}_0 \cup \{\infty\}$$

is given by

$$N_y(z, E) := \operatorname{card}(\{y^{-1}(z) \cap E\}). \tag{3.25}$$

Having this tools at hand, we can present a change of variables formula under minimal assumptions, which was given by Hajlasz [69]. The central idea is, that points hit multiple times by the transformation need to be taken into account as multiplicative factor.

Theorem 3.2.7 (Area formula)**.** *Let $y : \Omega \to \mathbb{R}^d$ be a mapping. If y is approximately differentiable almost everywhere, then y can be redefined on a zero set, such that the new y fulfills Lusin's condition. Furthermore the following statements hold for every measurable subset E and positive measurable function $u : \Omega \to \mathbb{R}$:*

(i) The functions $u(y)\det(\nabla y)$ and $u(z)N_y(z, \Omega)$ are measurable.

(ii) If moreover $u \geq 0$ then

$$\int_E u(y(x))\det(\nabla y(x))\mathrm{d}x = \int_{\mathbb{R}^m} u(z)N_y(z, \Omega)\mathrm{d}z. \tag{3.26}$$

(iii) If one of the functions $u(y)\det(\nabla y)$ and $u(z)N_y(z, E)$ is integrable then so is the other and the formula above holds.

Additionally we have

$$\int_\Omega u(x)\det(\nabla y(x))\mathrm{d}x = \int_{\mathbb{R}^d} \left(\sum_{w \in (y^{-1}(z) \cap \Omega)} u(w) \right) \mathrm{d}z. \tag{3.27}$$

Proof. See [69, Theorem 2] for the first part of the theorem.
The second part is given in [60, Chapter 1.2 Theorem 2] for Lipschitz mappings and can be generalized by following the proof by Hajlasz [69, Theorem 2]. □

A natural arising question is whether we are able to establish bounds for the Banach indicatrix, besides the fact that $N_f \in L^1$, whenever $\det(\nabla f) \in L^1$, which is directly deduced from the formula above. The following example shows that this might indeed be a difficult task:

Example 3.2.2 (Local versus global injectivity)**.** Let $\Omega = [0,1]^2$ and $n \in \mathbb{N}$. We define a sequence of functions $y_n : \Omega \to \mathbb{R}^2,\ n \in \mathbb{N}$ by setting for any $x = (x_1, x_2)$

$$y_n(x) = \begin{pmatrix} \mathrm{e}^{x_1} \cos(2\pi n x_2) \\ \mathrm{e}^{x_1} \sin(2\pi n x_2) \end{pmatrix} \tag{3.28}$$

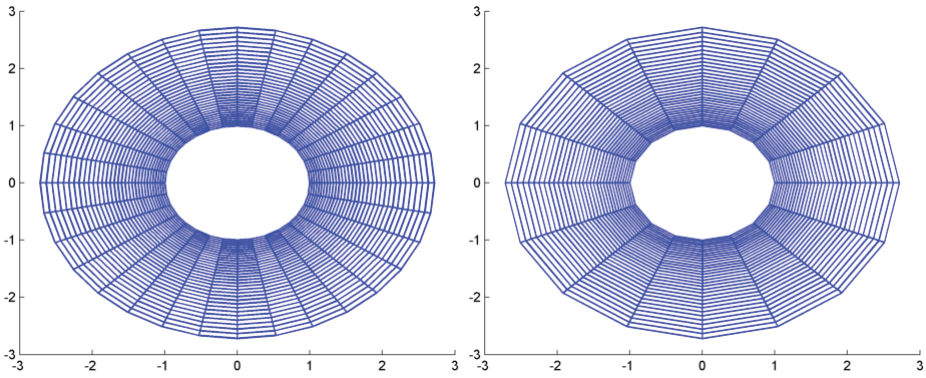


Figure 3.2: Visualization of the y_n (c.f (3.28)) for $n = 1, 2$. Note that the visible cells get bigger for $n = 2$, because the points in the ring are hit two times.

Obviously $y_n \in C^\infty(\Omega)\ \forall n \in \mathbb{N}$ with

$$\nabla y_n = \begin{pmatrix} \mathrm{e}^{x_1} \cos(2\pi n x_2) & -2\pi n\, \mathrm{e}^{x_1} \sin(2\pi n x_2) \\ \mathrm{e}^{x_1} \sin(2\pi n x_2) & 2\pi n\, \mathrm{e}^{x_1} \cos(2\pi n x_2) \end{pmatrix} \tag{3.29}$$

and thus:

$$\det(\nabla y_n)(x_1, x_2) = 2\pi n\, \mathrm{e}^{2x_1} > 0 \tag{3.30}$$

Therefore y_n is a local diffeomorphism for each $(x_1, x_2) \in [0,1]^2$. But also for each $(z_1, z_2) \in \mathrm{Im}(y_n) = B_\epsilon(0) \setminus B_1(0)$ there also exist n pairwise different tupels (x_1, x_2^i) with $f(x_1, x_2^i) = (y_1, y_2)$. It directly follows, that:

$$N_{y_n}(z, \Omega) = n \quad \forall z \in \mathrm{Im}(y_n) \tag{3.31}$$

Note that we presented this example for the sake of clarity in $\mathbb{R}^2$, but the extension into arbitrary dimensions is straightforward, as we illustrate for the three dimensional case:

$$y_n(x) = \begin{pmatrix} e^{x_1}\cos(2\pi n x_2) \\ e^{x_1}\sin(2\pi n x_2) \\ x_3 \end{pmatrix} \tag{3.32}$$

The L^1-norm of the gradient as well as the Jacobian determinant of this sequence of functions tends to infinity for growing n. Nevertheless this sequence shows that there might be difficulties in controling the Banach indicatrix in L^∞.
Furthermore this subject ist strongly linked to the topological degree, which was introduced by Browers in 1911 [21] and generalized to Sobolev mappings, by Giaquinta et al. [59]. We will not focus on this topic, but we will use some of this results in the remaining part of this thesis. We start by giving the generalized definition of the topological degree, see also[59]

Definition 3.2.8 (Topological degree)**.** Let $\Omega \subset \mathbb{R}^d$ be an open set. For an almost everywhere approximately differentiable map y with Jacobian Dy the degree of y is defined as

$$\deg(y, \Omega, z) := \sum_{x \in v^{-1}(z)} \operatorname{sgn}(\det(Dv(x))) \tag{3.33}$$

Remarks and Extensions. The topological degree is strongly related to the Banach indicatrix, since it coincides with the Banach indicatrix for certain mappings. As a direct consequence of [58, Chapter 1, Proposition 2] we obtain for orientation preserving mappings:

$$\deg(u, \Omega, \cdot) = N_u(\cdot, \Omega) \qquad \text{a.e.} \tag{3.34}$$

An interesting property of the topological degree is that it is completely determined on the boundary for sufficiently regular functions. This is phrased in the following proposition:

Proposition 3.2.9. *Let Ω be a bounded Lipschitz domain in $\mathbb{R}^d$ and let v_1, v_2 be mappings in $W^{1,d-1}(\Omega, \mathbb{R}^d)$ with $\operatorname{cof}(\nabla v_i) \in L^{\frac{d}{d-1}}$. Suppose that*

$$v_1 = v_2 \qquad \text{on} \qquad \partial\Omega \tag{3.35}$$

in the sense of $W^{1,d-1}$ traces. Then

$$\deg(v_1, \Omega, y) = \deg(v_2, \Omega, y) \tag{3.36}$$

Proof. See [58], Chapter 2, Proposition 1. □

To conclude this brief summary, we mention that there is indeed a specification to Sobolev functions, which are injective. This leads to the field of the so called weak diffeomorphisms, which we discuss shortly in the next section.

3.3 Weak Diffeomorphisms

We will present some basic ideas and Theorems from [58], how to generalize diffeomorphisms for Sobolev mappings, which are not necessarily differentiable. Giaquinta's central idea for a mapping

$$y : \Omega \subset \mathbb{R}^d \to \mathbb{R}^m \tag{3.37}$$

is to use properties of the graph $\mathcal{G} \subset \mathbb{R}^d \times \mathbb{R}^m$. Based on this idea he introduced so called Cartesian currents [56]. As the description of these currents is beyond the scope of this thesis, we will only present the results from [58] related to weak diffeomorphisms and refer to [58, 60, 62] for a detailed course on Cartesian currents. Because we do not want to carry out too much of the theory on Cartesian currents, we define a class of transformations, which contains transformations fulfilling some rather complicated requirements from the theory of Cartesian currents.

Definition 3.3.1. We define the following two subclasses of Sobolev mappings:

$$\begin{aligned} A_{p,q}(\Omega) :=& \{y \in W^{1,p}(\Omega;\mathbb{R}^m) \mid \operatorname{cof}(\nabla y) \in L^q\}, \\ A^+_{p,q}(\Omega) :=& \{y \in A_{p,q}(\Omega) \mid \det(\nabla y) > 0 \quad \text{a.e.}\}. \end{aligned}$$

Next we introduce weak inverses as in [58]:

Definition 3.3.2 (Weak inverse)**.** Given a measurable map

$$y : \Omega \to \hat{\Omega} \qquad \lambda(\Omega) > 0 \quad \lambda(\hat{\Omega}) > 0. \tag{3.38}$$

We say that

1. y is weakly invertible with weak inverse $\hat{y}$, if and only if

$$\hat{y}(y(x)) = x \qquad \text{for almost every } x \in \Omega, \tag{3.39}$$
$$y(\hat{y}(z)) = z \qquad \text{for almost every } z \in \hat{\Omega}. \tag{3.40}$$

2. y is a weak one-to-one transformation, iff there exists a measurable map

$$\hat{y} : \hat{\Omega} \to \Omega,$$

such that

a) y and $\hat{y}$ fulfill Lusin's condition (3.23).

b) y and $\hat{y}$ are the inverses of the respective other.

The next theorem gives us some properties of the inverse of a mapping:

Theorem 3.3.3. *Let $y : \Omega \to \hat{\Omega}$ be a weakly invertible map with inverse $\hat{y}$. Suppose that*

(i) y satisfies Lusin's condition (3.23),

(ii) y is almost everywhere approximately differentiable in Ω.

Then $\hat{y}$ is approximately differentiable almost everywhere in $\hat{\Omega}$. Moreover:

$$Dy(\hat{y}(z))D\hat{y}(z) = Id_{\hat{\Omega}} \qquad \text{for a.e.} \quad z \in \hat{\Omega} \tag{3.41}$$

$$D\hat{y}(y(x))Dy(x) = Id_{\Omega} \qquad \text{for a.e.} \quad x \in \Omega \tag{3.42}$$

Proof. See [58, Chapter 3, Theorem 2]. □

With this at hand Giaquinta et al. [58] define global invertibility by using properties of the graph of a map by

Definition 3.3.4 (Global invertibility for a.e. approximately differentiable mappings)**.** Let y be an a.e. approximately differentiable map from $\Omega \subset \mathbb{R}^d$ into $\mathbb{R}^m$ with $\det(Dy) \in L^1(\Omega)$ satisfying

$$\det(Dy) > 0 \qquad \text{a.e. in } \Omega. \tag{3.43}$$

We say y is globally invertible if and only if

$$\int_{\Omega} \phi(x, y(x)) \det(Dy(x)) \mathrm{d}x \leq \int_{\mathbb{R}^m} \left(\sup_{x \in \Omega} \phi(x, z) \right) \mathrm{d}z \tag{3.44}$$

holds for all $\phi \in C_c^0(\Omega \times \mathbb{R}^m)$ with $\phi \geq 0$.

The definition above is related to the area formula; for any function $\phi \in C_c^0(\Omega \times \mathbb{R}^m)$ we observe:

$$\int_{\Omega} \phi(x, y(x)) \det(Dy)(x) \mathrm{d}x \leq \int_{\Omega} \underbrace{\sup_{x \in \Omega} \phi(x, y(x))}_{=:\psi(y(x))} \det(Dy(x)) \mathrm{d}x = \int_{\mathbb{R}^m} \sup_{x \in \Omega} \phi(x, z) N_y(z, \Omega) \mathrm{d}z$$

This illustrates, how the invertibility condition (3.44) can be violated by functions which are not injective on sets with positive measure. According to [58] global invertibility can be defined equivalently in several other ways:

Proposition 3.3.5. *Let $y : \Omega \subset \mathbb{R}^d \to \mathbb{R}^m$ be a.e. approximately differentiable in Ω with $\det(Dy) \in L^1(\Omega)$ and $\det(Dy) \geq 0$ a.e. in Ω. Then the following claims are equivalent:*

(i) y is globally invertible.

(ii) For any $\phi \in C_c^0(\mathbb{R}^m)$ with $\phi \geq 0$ y satisfies the inequality

$$\int_{\Omega} \phi(y(x)) \det(Dy)(x) \mathrm{d}x \leq \int_{\mathbb{R}^m} \phi(z) \mathrm{d}z. \tag{3.45}$$

(iii) For almost every $z \in \mathbb{R}^m$ we have

$$N(y, \Omega, z) \leq 1. \tag{3.46}$$

(iv) For almost every $z \in \mathbb{R}^m$ we have

$$N_y(\Omega, z) = \chi_{y(\Omega)}(z) := \begin{cases} 1 & z \in y(\Omega) \\ 0 & z \notin y(\Omega) \end{cases}. \tag{3.47}$$

(v) We have

$$\int_\Omega \det(Dy)(x)\mathrm{d}x = \mathscr{H}^m(\tilde{y}(\Omega)) \tag{3.48}$$

where $\mathscr{H}^m$ is the m-dimensional Hausdorff measure[2] and $\tilde{y}$ a Lusin representative.

(vi) The inequality

$$\int_\Omega \det(Dy)(x)\mathrm{d}x \leq \mathscr{H}^m(y(\Omega)) \tag{3.49}$$

holds for any representative of y.

Proof. See [58, Chapter 5, Proposition 1] □

Having this in mind we can now define a norm for the class of weak diffeomorphisms:

Definition 3.3.6 (Norm for weak diffeomorphisms)**.** For any almost everywhere approximately differentiable map $y : \Omega \subset \mathbb{R}^d \to \mathbb{R}^m$ we set

$$|M(Dy)| := \left(1 + |Dy|^2 + |\operatorname{cof}(Dy)|^2 + |\det(Dy)|^2\right)^{\frac{1}{2}} \tag{3.50}$$

and define

$$\|y\|_{\mathrm{dif}^{p,q}} := \int_\Omega \left(|y|^p + |M(Dy)|^p + \frac{|M(Dy)|^q}{|\det(Dy)|^{q-1}}\right) \mathrm{d}x. \tag{3.51}$$

Now we can define the class of weak diffeomorphisms:

Definition 3.3.7 (Space of weak diffeomorphisms)**.** We say that a map $y : \Omega \subset \mathbb{R}^d \to \mathbb{R}^m$ belongs to the class $\tilde{\mathrm{dif}}^{p,q}(\Omega, \mathbb{R}^m)$ for $p, q \geq 1$, if and only if:

1. $|M(Dy)| \in L^p$.
2. y has a closed graph in $\Omega \times \mathbb{R}^m$.

[2] Hausdorff measures provide s-dimensional measures in $\mathbb{R}^m$ for any $0 \leq s \leq m$. For Euclidean vector spaces the Hausdorff measure $\mathscr{H}^m$ coincides with the Lebesgue measure. See [60, 47] for more details.

3. $\det(Dy) > 0$ a.e. in Ω.

4. y is globally invertible.

5. $\|y\|_{\mathrm{dif}^{p,q}} < \infty$.

A map only fulfilling the first three conditions is called a weak local diffeomorphism.

Remarks and Extensions. Since giving a thorough definition of the second property requires insight in the theory of Cartesian currents, we will not elaborate further on this subject. Details can be found e.g. in [56, 60, 62]. However we can state that $y \in A_{d-1,\frac{d}{d-1}}$ is sufficient [59] but not necessary [61] to guarantee the closedness of the graph.

A weak diffeomorphism y can be expressed as the limit of a sequence y_n of orientation preserving C^1 diffeomorphisms [56]. Furthermore closeness and compactness results for this class can be stated ([58, Chapter 5, Theorems 3 and 4]). This relies heavily on weak convergence in $\tilde{\mathrm{dif}}^{p,q}$, which again needs a deep insight into the theory of currents. Thus we present a similar result published by Henao and Mora-Corral [75] instead.

Theorem 3.3.8 (Injectivity as closed constraint)**.** *For each $j \in \mathbb{N}$ let $y_j, u : \Omega \subset \mathbb{R}^d \to \mathbb{R}^m$ be a.e. approximately differentiable. Assume furthermore, that*

$$y_j \in W^{1,p}(\Omega, \mathbb{R}^m) \quad p \geq d-1 \qquad \det(Dy_j) \in L^1(\Omega) \tag{3.52}$$

as well as

$$\operatorname{cof}(Dy_j) \in L^g(\Omega) \quad q \geq \frac{p}{p-1} \qquad \sup_{j\in\mathbb{N}} \|\operatorname{cof}(Dy_j)\|_1 < \infty. \tag{3.53}$$

Suppose that there exists $\theta \in L^1(\Omega)$ such that $\theta > 0$ a.e. and

$$y_j \to y \qquad \det(Dy_j) \rightharpoonup {}^{3}\theta \text{ in } L^1(\Omega) \tag{3.54}$$

as $j \to \infty$. Assume that for each $j \in \mathbb{N}$ the function y_j is one-to-one a.e. with $\det(Dy_j) > 0$ a.e.. Then

(i) $\theta = |\det(Dy)|$ a.e.,

(ii) y is one-to-one a.e..

Proof. Because of (3.53) we can deduce by [75, Proposition 2] that the functional $\overline{\mathcal{E}}$ introduced by Henao and Mora-Corral is zero for every y_j. Now the assertion follows directly by [75, Theorem 2]. □

[3] This denote weak convergence; see Definition 3.4.3.

3.4 The Direct Method in the Calculus of Variations

The goal of this section is to derive a method for showing the existence of minimizers of maps with functions as arguments. While the basic principles of functional analysis we present can e.g. be found in [119], an introduction to the direct method in the calculus of variations can be found in [39]. We will define this class of mappings mapping into the real numbers:

Definition 3.4.1 ((Linear) functional)**.** Let $\mathcal{X}$ be a $\mathbb{R}$-vector space . A map

$$J : \mathcal{X} \to \mathbb{R} \tag{3.55}$$

is called a functional.
If J additionally fulfills

$$J(\alpha x_1 + \beta x_2) = \alpha J(x_1) + \beta J(x_2) \qquad \forall x_1, x_2 \in \mathcal{X}, \alpha, \beta \in \mathbb{R} \tag{3.56}$$

then J is called a linear functional.

The space of linear, continuous functionals plays an important role in defining weaker forms of convergence.

Definition 3.4.2 (Dual space)**.** Let $\mathcal{X}$ be a vector space. The dual space is given by

$$\mathcal{X}^* := \{J : \mathcal{X} \to \mathbb{R} \quad | \quad J \text{ continuous and linear}\}. \tag{3.57}$$

We will use the dual space to define some different types of convergence. To weaken the requirements of convergence in the norm one just investigates whether applying linear functionals to a sequence leads to convergence.

Definition 3.4.3 (Weak convergence)**.** Let x_n be a sequence in $\mathcal{X}$ and $x \in \mathcal{X}$. We say x_n converges weakly to x, iff

$$\phi(x_n) \to \phi(x) \qquad \forall \phi \in \mathcal{X}^* \tag{3.58}$$

and use the notation

$$x_n \rightharpoonup x.$$

With the same idea one is able to define a weak form of convergence for linear functionals.

Definition 3.4.4 (Weak-star convergence)**.** Let $\mathcal{X} = Z^*$ be the dual of a normed vector space, x_n be a sequence in $\mathcal{X}$ and $x \in \mathcal{X}$. We say x_n converges weakly-star to x, iff

$$x_n(z) \to x(z) \qquad \forall z \in Z \tag{3.59}$$

and use the notation

$$x_n \rightharpoonup^* x.$$

Having defined these basics, we turn our focus on how to show the existence of minimizers for problems of the form

$$\inf_{\rho} J(\rho). \tag{3.60}$$

Since we end up with minimization problems in our modeling for image and motion estimation we present the standard approach to show the existence of minimizers in the calculus of variations. We need to verify two conditions, in order to show existence of a minimizer. The first one is a weaker version of a continuity property:

Definition 3.4.5 (Lower semicontinuity)**.** Let $J : \mathcal{X} \to \mathbb{R}$ be a functional. J is lower semicontinuous, iff for every sequence $x_n \to x$

$$J(x) \leq \liminf_{n\to\infty} J(x_n). \tag{3.61}$$

With that at hand, we can give the following theorem, which inspires the direct method in the calculus of variations.

Theorem 3.4.6 (Fundamental theorem of optimization)**.** *Let $J : \mathcal{X} \to \mathbb{R}$ be a functional. Let J fulfill the following two conditions:*

1. *There is a $c \in \mathbb{R}$, such that the sublevel set $\{x \in \mathcal{X} | J(x) \leq c\}$ is compact and non-empty.*
2. *J is lower semicontinuous.*

Then J has a minimizer.

Proof. Let x_n be an infimizing sequence of the functional, then there exists $N \in \mathbb{N}$, such that $J(x_n) \leq c \; \forall n \geq N$. Since this sublevel set is compact, we find a convergent subsequence $\tilde{x}_n$. For the limit $\tilde{x}$ of this subsequence we obtain by the lower semicontinuity of J:

$$J(\tilde{x}) \leq \liminf_{n\to\infty} J(\tilde{x}_n) = \inf J. \tag{3.62}$$

Thus $\tilde{x}$ is a minimizer. □

Verifying the compactness of sublevel sets might turn out to be a difficult task in practical applications. While in finite dimensions boundedness of a closed set is equivalent to compactness, unfortunately this does in general not hold for infinite dimensional spaces. However with the famous theorem of Banach-Alaoglu we can prove a similiar result for the weak-star topology:

Theorem 3.4.7 (Banach-Alaoglu)**.** *Let $\mathcal{X} = Z^*$ be the dual of a Banach space Z and M a bounded set in $\mathcal{X}$. Then M is precompact in the weak-star topology.*

Proof. See for example [37], Theorem 3.1 □

The boundedness can be shown by the so called coercivity property:

Definition 3.4.8 (Coercivity)**.** Let $J : \mathcal{X} \to \mathbb{R}$ be a functional. J is called coercive, iff

$$J(x) \to \infty \qquad \text{whenever} \qquad \|x\| \to \infty.$$

Having this tool at hand, in the remainder of this thesis we will only show a coercivity property of an examined functional, because we can deduce the compactness of sub level sets the by Banach-Alaoglu theorem 3.4.7.
There are different approaches, when the direct method fails, such as [30, 83] . The most common way is to take a probabilistic approach, where a so called Young measure gives the probability that a functional has its weak limit as value for each x. These Young measures were first presented as generalized curves by L.C Young in 1937 [148]. A detailed description of Young measures can be found for example in [107, 83].

3.5 Theory of (Linear) Operators

Since many problems in mathematical imaging can be stated with the help of linear operators we give a short overview of the most important definitions and theorems on this subject. A more detailed description of this subject as well as the definitions and results we present in the remainder of this section can be found in most courses about functional analysis, such as [4, 146, 37].
We start with the definition of a linear operator:

Definition 3.5.1 (Linear operator)**.** Let $\mathcal{X}, \mathcal{Y}$ be vector spaces; a map $T : \mathcal{X} \to \mathcal{Y}$ is called a linear operator, iff

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y) \qquad \forall x, y \in \mathcal{X}, \alpha, \beta \in \mathbb{R}. \tag{3.63}$$

The following definitions captures some notations used in the remainder of this thesis.

Definition 3.5.2. For any linear operator $T : \mathcal{X} \to \mathcal{Y}$ between Banach spaces we denote:

- The domain $\mathcal{D}(T)$ as the set of points $x \in \mathcal{X}$, such that Tx is defined.
- The kernel $\mathcal{N}(\mathcal{T})$ is defined as $\{x \in \mathcal{X} \,|\, Tx = 0\}$.
- The range $\mathcal{R}(\mathcal{T})$ is defined as $\{y \in \mathcal{Y} \,|\, \exists x \in \mathcal{X} \quad \text{with} \quad Tx = y\}$.

For any subspace $\mathcal{X}_0$ of $\mathcal{X}$ the orthogonal complement is defined by

$$\mathcal{X}_0^{\perp} := \{v \in \mathcal{X}_0^* \,|v(x) = 0 \quad \forall x \in \mathcal{X}_0\}$$

In fact continuity of linear operators can be characterized in many equivalent ways:

Theorem 3.5.3 (Continuity for linear operators)**.** *Let $T : \mathcal{X} \to \mathcal{Y}$ be a linear operator. Then the following claims are equivalent:*

(i) *T is continuous.*

(ii) *There is a $c \geq 0$, such that: $\|Tx\|_{\mathcal{Y}} \leq c\|x\|_{\mathcal{X}}$ for all $x \in \mathcal{X}$.*

(iii) *There is $c \geq 0$ such that $\|Tx\|_{\mathcal{Y}} \leq c$ for all $x \in B_1(0)$.*

The next step is now to define the space of linear operators and equip it with a norm:

Definition 3.5.4 (Space of linear operators and operator norm)**.** Let $\mathcal{X}, \mathcal{Y}$ be vector spaces. Then the vector space of linear operators mapping from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $\mathcal{L}(\mathcal{X}, \mathcal{Y})$. This vector space is equipped with the following so called operator norm

$$\|T\| := \inf\{c \geq 0 \,|\, \|Tx\|_{\mathcal{Y}} \leq c\|x\|_{\mathcal{X}} \forall x \in \mathcal{X}\}. \tag{3.64}$$

Remarks and Extensions. If $\mathcal{X}$ is a normed vector space and $\mathcal{Y}$ complete, then $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is a Banach space.

The next theorem shows how an operator on a given dense subset can be extended to the whole space:

Theorem 3.5.5 (Unique continuous extension)**.** *Let $\mathcal{X}$ be a normed vector space, $\mathcal{X}_0$ a dense normed subspace. For a Banach space $\mathcal{Y}$ let $T_0 \in \mathcal{L}(\mathcal{X}_0, \mathcal{Y})$. Then there is exactly one operator $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, such that*

$$T_0 x = Tx \qquad \forall x \in \mathcal{X}_0. \tag{3.65}$$

Additional there is

$$\|T_0\| = \|T\|. \tag{3.66}$$

Besides continuity, there is another important definition we will introduce shortly in the following:

Definition 3.5.6 (Adjoint operator)**.** Let $\mathcal{X}, \mathcal{Y}$ be Hilbert spaces and $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. Iff T^* fulfills

$$\langle Tx, y\rangle_{\mathcal{Y}} = \langle x, T^* y\rangle_{\mathcal{X}} \qquad \forall x \in \mathcal{X}, y \in \mathcal{Y} \tag{3.67}$$

then T^* is called the adjoint operator of T.

Remarks and Extensions. The existence of the adjoint operator follows from Riesz representation theorem [37, Thm. 3.4]. Furthermore the concept of adjoints can also be generalized to Banach spaces [64].

Having stated the main definitions and theorems for linear operators we will draw our attention on a weaker form of operator convergence:

Definition 3.5.7 (Strong operator topology)**.** Let $T_n, T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. We say the sequence T_n converges to T in the sense of the strong operator topology, and write

$$T_n \xrightarrow{s} T \tag{3.68}$$

iff

$$\lim_{n\to\infty} T_n x = Tx \qquad \forall x \in \mathcal{X}. \tag{3.69}$$

An useful tool to prove convergence in the sense of the strong operator convergence is provided by the next theorem, which allows us to verify this only on a dense subspace:

Theorem 3.5.8. *Let $T_n, T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, such that*

$$\sup \|T_n\| < \infty. \tag{3.70}$$

Then the following statements are equivalent:

(i) There is a dense subspace $\mathcal{X}_0$ of $\mathcal{X}$, such that $\lim_{n\to\infty} T_n x = Tx$ for all $x \in \mathcal{X}_0$.

(ii) T_n converges to T in the sense of the strong operator topology.

To conclude the remarks about strong operator convergence we make an observation concerning the product of convergent sequences:

Theorem 3.5.9. *Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be Banach spaces, $T_n, T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, $S_n, S \in \mathcal{L}(\mathcal{Y}, \mathcal{Z})$, such that*

$$T_n \xrightarrow{s} T \qquad \text{and} \qquad S_n \xrightarrow{s} S \tag{3.71}$$

in the sense of the strong operator topology. Then

$$S_n T_n \xrightarrow{s} ST \tag{3.72}$$

again in the sense of the strong operator topology.

After having presented the most important results about the strong operator topology we mention that there is also a weaker one:

Definition 3.5.10 (Weak operator topology)**.** Let $T_n, T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. T_n converges to T in the sense of the weak operator topology, iff

$$\lim_{n\to\infty} \varphi(T_n x) = \varphi(Tx) \qquad \forall x \in \mathcal{X}, \varphi \in \mathcal{Y}^*. \tag{3.73}$$

This concludes our short summary about linear operators. In the following we will focus on an interesting special case, namely compact operators.

Compact Operators

Compact operators have many applications in mathematical imaging: The projection process in indirect imaging is for example often modeled with the help of compact operators. In this subsection we will focus on the definition and some basic properties of compact operators. We follow Conways presentation [37]; all proofs for the theorems presented in this subsection can be found there.

Definition 3.5.11 (Compact operator)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $K : \mathcal{X} \to \mathcal{Y}$ a linear operator. We call K a compact operator, iff the closure $\mathrm{cl}(K(B_1(0)))$ is compact.

Another property compact operators often possess is the following:

Definition 3.5.12 (Complete continuity)**.** Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces. A linear operator T is called completely continuous, iff:

$$x_n \rightharpoonup x \qquad \Rightarrow \qquad Ax_n \to Ax. \tag{3.74}$$

The next proposition states, under which circumstances those two properties are equivalent:

Proposition 3.5.13. *Let $\mathcal{X}, \mathcal{Y}$ be Banach spaces and $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$:*

(a) If T is compact, then T is completely continuous.

(b) If $\mathcal{X}$ is reflexive and T is completely continuous, then T is compact.

The next theorem gives another criterion for the compactness of a linear operator:

Theorem 3.5.14 (Schauders theorem)**.** *Let $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$, then T is compact iff T^* is compact.*

We have now collected basic definitions and presented the theorems we will use as tools in the remainder of this thesis. A much more detailed overview on this subject is given in the references at the start of this section.

4 Inverse Problems in Imaging

In this chapter we describe inverse problems occuring in mathematical imaging. First we describe the field of inverse problems in general. We start with a rather general introduction to inverse and ill-posed problems by giving some basic definitions. Again this chapter only serves as an overview on definitions and methods used in the literature. After presenting some general ideas for computing meaningful solutions for ill-posed problems, we turn our focus on reconstruction and motion estimation, which is the major focus in this thesis.
Following the introduction by Bertero and Boccacci [16] we present Kellers [82] description of inverse problems:

Inverse Problems

We call two problems **inverses** of one another if the formulation of each involves all or part of the solution of the other. Often, for historical reasons, one of the two problems has been studied extensively for some time, while the other is newer and not so well understood. In such cases the former is called the **direct problem**, while the latter one is called the **inverse problem**.

As an example of two inverse problems Keller gives the problem of finding the roots of an polynomial with order n as the direct problem and finding a polynomial with order n with given roots as the inverse problem. Clearly in this case the inverse problem is easier to solve but also the solution is not unique, since one can multiply any given solution by a constant $c \neq 0$. In contrast, finding the roots of a given polynomial might cause some effort, nevertheless the solution is unique (if we take multiple roots into account). This behaviour is formulated in the following definition by Hadamard [68]:

Definition 4.0.1 (Well and ill-posed problems)**.** According to Hadamards definition a problem is **well-posed** if the following three conditions hold.

- A solution exists.
- The solution is unique.
- The solution's behaviour changes continuously with the initial conditions.

If one of the three conditions is not satisfied, the problem is called **ill-posed**.

Having formulated the main topic of this chapter, we will first present some ways to alleviate difficulties in dealing with ill-posed problems and then lie our focus on reconstruction and motion estimation problems. These two tasks are frequently occurring in classical imaging and furthermore these are the major problems investigated throughout this thesis.

4.1 Dealing with Ill-Posed Problems

In this section we will describe some common approaches to solve the problem

$$Ku = f, \tag{4.1}$$

where K is an operator between Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. Typically, only noisy data is available, thus we have in fact to solve the problem

$$Ku = f^\delta \tag{4.2}$$

for noisy data f^δ with

$$\|f - f^\delta\| \leq \delta. \tag{4.3}$$

A way to define an inverse for the operator K is given by [44]

Definition 4.1.1 (Moore-Penrose inverse)**.** The Moore-Penrose (generalized) inverse $K^\dagger$ of $K \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ is defined as the unique linear extension of $\tilde{K}^{-1}$ to

$$\mathcal{D}(K^\dagger) := \mathcal{R}(K) + \mathcal{R}(K)^\perp \tag{4.4}$$

with

$$\mathcal{N}(K^\dagger) = \mathcal{R}(K)^\perp \tag{4.5}$$

where

$$\tilde{K} := K|_{\mathcal{N}(K)^\perp} : \mathcal{N}(K)^\perp \to \mathcal{R}(K)$$

According to [44] this inverse is well-defined.
Typically problem (4.1) is an ill-posed, inverse problem in the sense that the Moore-Penrose inverse $K^\dagger$ for the forward operator K is not continuous.
It is even possible, that (for example due to modeling errors) only a noisy version $\tilde{K}$ of the forward operator is known. In the linear, discrete sense K is an ill-conditioned matrix. Thus simply applying the Moore-Penrose inverse $K^\dagger$ to the noisy data y^δ will in general not lead to meaningful solutions. We will now describe regularization and statistical methods, designed to deal with ill-posed problems.

4.1.1 Regularization Operators

Following the introduction to regularization from [44] we give the central idea of regularization:

The Central Idea of Regularization

In general terms, **regularization** is the approximation of an ill-posed problem by a family of well-posed problems.

This leads to the definition of regularization:

Definition 4.1.2 (Regularization operator)**.** Let $K : \mathcal{U} \to \mathcal{Y}$ be a bounded linear operator between the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, $\alpha_0 \in (0, \infty]$. For every $\alpha \in (0, \alpha_0)$ let

$$R_\alpha : \mathcal{Y} \to \mathcal{X}$$

be a continuous (not necessarily linear) operator. The family $\{R_\alpha\}$ is called a regularization or a regularization operator (for $K^\dagger$), iff for all $f \in \mathcal{D}(K^\dagger)$ there exists a parameter choice rule $\alpha = \alpha(\delta, f^\delta)$ such that

$$\limsup_{\delta \to 0} \{ \| R_{\alpha(\delta, f^\delta)} f^\delta - K^\dagger f \| \quad | \quad f^\delta \in \mathcal{Y}, \| f - f^\delta \| \leq \delta \} = 0 \tag{4.6}$$

holds. Here

$$\alpha : \mathbb{R}^+ \times \mathcal{Y} \to (0, \alpha_0)$$

is such that

$$\limsup_{\delta \to 0} \{ \alpha(\delta, f^\delta) | f^\delta \in \mathcal{Y}, \| f^\delta - f \| \leq \delta \} = 0. \tag{4.7}$$

For a specific $f \in \mathcal{D}(K^\dagger)$ a pair (R_α, α) is called a (convergent) regularization method (for solving $Ku = f$) if (4.6) and (4.7) hold.

Remarks and Extensions. This definition can be extended to nonlinear problems in a straightforward way, but we choose to give the original one from [44]. Furthermore we can also deal with noisy versions $\tilde{K}$ of the operator with $\| \tilde{K} - K \| \leq \eta$; then the parameter choice rule would depend on δ, η, y^δ and $\tilde{K}$.

An extensive overview of different regularization operators and their numerical implementation can e.g. be found in [71, 72]. Before we present examples for regularization operators, we turn our focus towards parameter choice rules.

Parameter Choice Rules

Engl et al. [44] distinguish between two different parameter choice rules:

Definition 4.1.3 (Parameter choice rules)**.** Let α be a parameter choice rule according to Definition 4.1.2. If α does not depend on y^δ but only on δ, then we call α an a priori parameter choice rule and write $\alpha = \alpha(\delta)$. Otherwise we call α an a posteriori parameter choice rule.

Note that according to Bakushinskii [9] parameter choice rules depending only on the data, but not on the noise level δ, cannot lead to convergent regularization methods.
As we saw before the desired properties of regularization operators depend on the choice of the regularization parameter α. In praxis this might be an issue, since the noise level δ is not known in general and needs to be estimated too. The general idea is to measure the difference between the projected solution of the regularization operator and the data. If this difference is smaller than the noise level δ, we cannot expect a meaningful solution, since mostly the noise is matched. Morozov [102] suggested to pick parameters fulfilling a discrepancy principle.

Definition 4.1.4 (Morozov Discrepancy principle)**.** Let an inverse (and ill-posed) problem of the form (4.1) be given. We say that a regularization parameter α fulfills the Morozov discrepancy priniciple for $1 \leq \tau_1 < \tau$, provided

$$\tau_1 \delta \leq \|KR_\alpha(f^\delta) - f^\delta\| \leq \tau_2. \tag{4.8}$$

We will not elaborate further on this matter, but refer to [120] for an evaluation of the performance for different parameter choice rules and to [43] for a discussion about discrepancy principles.
Having stated this common parameter choice strategies we conclude this section by mentioning that estimating the noisy level from given data is a difficult task (see for example [81, Chapter 2, Example 4]).

(Truncated) Singular Value Decomposition

A basic tool to construct regularization operators is the singular system - sometimes called singular value expansion - [73] $(\sigma_n; v_n, u_n)$, which can be found for a compact operator K, such that

$$Kx = \sum_{n=1}^{\infty} \sigma_n \langle x, v_n \rangle u_n. \tag{4.9}$$

The Moore-Penrose inverse of the operator K can be expressed with help of the system (4.9) via

$$K^\dagger y = \sum_{n=1}^{\infty} \frac{\langle y, u_n \rangle}{\sigma_n} v_n. \tag{4.10}$$

This expression of the inverse operator shows that, errors in the data might be blown up by the small singular values. Inspired by this observation the idea of the truncated singular value decomposition is to stop the summation at an index J, where the singular values are smaller than a specified bound α. Thus we can define an operator R_J via

$$R_J y = \sum_{n=1}^{J} \frac{\langle y, u_n \rangle}{\sigma_n} v_n. \tag{4.11}$$

This finite summation is called **truncad singular value decomopsition** and is often used as an regularization method. An advantage of this method is, that the (truncated) SVD only needs to be computed once for a specific operator and thus the truncated SVD solution can be computed fast for different data. Unfortunately we are not directly able to incorporate a priori knowledge of the desired solution into the regularization process. See for example [70, 73] for more details on the use of this method as regularization operator.

Tikhonov Regularization

Tikhonov-Regularization, also known as **Tikhonov-Phillips regularization** is a widely used regularization method. Inspired by the idea to enforce smoothness assumptions on the solution, it was designed by Tikhonov [143, 144], and slightly later by David L. Phillips [115]. We will present now shortly the central ideas, which can also be found - with much more details - in Chapter 5 of [44].
For the standard inverse (and perhaps ill-posed problem) (4.1) the normal equation is given by

$$K^*Ku = K^*y. \tag{4.12}$$

The solutions of this equation are the least squares solutions of problem (4.1) (compare e.g. [44, Theorem 2.6]). For any $\alpha > 0$ we consider

$$K^*Ku + \alpha u = K^*y^\delta. \tag{4.13}$$

Solving this equation is called Tikhonov regularization. Obviously this equation has the solution

$$\hat{u} = (K^*K + \alpha I)^{-1}K^*y^\delta \tag{4.14}$$

which can also be expressed with the singular system as

$$\hat{u} = \sum_{n=1}^{\infty} \frac{\sigma_n}{\sigma_n^2 + \alpha}\langle y^\delta, u_n\rangle v_n. \tag{4.15}$$

This representation shows that - similar to the truncated SVD - errors in the data are not blown up as σ_n tends to zero, since $\frac{\sigma_n}{\sigma_n^2+\alpha}$ is bounded for $n \to \infty$ and therefore $\sigma_n \to 0$. The solution $\hat{u}$ can also be characterized variationally:

$$\hat{u} := \min_u J(u) := \|Ku - y^\delta\| + \alpha\|u\|^2 \tag{4.16}$$

Note that we can choose different norms in (4.16) as well as replace the identity in (4.13) by a so called Tikhonov matrix Γ [121]. In contrast to the formulation with the normal equation (4.13) the functional (4.16) also allows the use of nonlinear operators. We complete this section by mentioning that for the Tikhonov regularization is in fact a regularization operator in the sense of Definition 4.1.2 [44, Chapter 5, Theorem 5.2].
An advantage of Tikhonov regularization over the truncated SVD is that, the Tikhonov Matrix Γ allows us to incorporate a priori knowledge on the desired solution be choosing an inappropriate norm. Unfortunately we are not able to incorporate additional knowledge on the characteristic of the noise corruption in the data into this process, thus we turn our focus to statistical methods, which are based on incorporating this information.

4.1.2 Statistical Methods

Since noise corrupting the data in many inverse problems has a random nature [81], another approach is to take the properties of this noise occurrence into account in the

model. In this section we will describe two common approaches to incorporate the process of corrupting data with random noise into specially designed strategies to solve ill-posed problems suffering from noise. The first idea is to interpret the noisy data f^δ as the realization of a random variable with the true data as mean value, and to search for a solution x, which has the biggest probability of producing this data. By contrast, in Bayesian methods not only the data but also the solution is regarded as the realization of a random variable. We will describe both approaches and draw a link between Bayesian methods and Tikhonov regularization. For a detailed review on statistical methods in inverse problems we refer to [79, 81], where the described contents can be found in much greater detail. Technically a probability is a measure on a σ-algebra, which assigns a probability to all events collected in the σ-algebra [12]. In the following we will not focus on this probability triple, but just use an appropriate probability p. A detailed discussion on this matter is given in [79, Chapters 3,7].

Maximum-Likelihood Estimation

The central idea of the Maximum-Likelihood estimation is to pick the solution $\hat{x}$, which maximizes the likelihood of measuring the data f^δ, such that the following equation

$$p(f^\delta|\hat{x}) \geq p(f^\delta|x) \qquad \forall x \tag{4.17}$$

holds for any estimator x [79]. We take a short look into the computation of these estimators: We assume that the data vector $f^\delta \in \mathbb{R}^N$ is a collection of N independent identically distributed random variables $F = (F_i)_i$. So the likelihood of measuring the actual realization f^δ is given by

$$L(F, F) = \prod_{i=1}^{N} p(F_i = f_i^\delta). \tag{4.18}$$

To maximize this product we apply the logarithm, which makes (4.18) a sum, the so called log-likelihood

$$\log\left(\prod_{i=1}^{N} p(Y_i = y_i^\delta)\right) = \sum_{i=1}^{N} \log(p(Y_i = y_i^\delta)). \tag{4.19}$$

To obtain a minimization problem, we minimize the negative log-likelihood, instead of the maximization. The following example highlights possible difficulties in this approach (compare [81, Chapter 5.1.1]):

Example 4.1.1 (Maximum-Likelihood Estimation for Gaussian Noise)**.** Let f^δ be a realization of a random variable Y with

$$Y = Ax + N \tag{4.20}$$

where N is $(0, \sigma^2)$ normal distributed and A is an (ill conditioned) measurement matrix. The likelihood of observing f^δ is given by

$$p(f^\delta|x) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{(Ax)_i - f_i^\delta}{\sigma}\right)^2\right). \tag{4.21}$$

By applying the logarithm we obtain:

$$\log(p(f^\delta|x)) = \sum_{i=1}^{N} \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{(Ax)_i - f_i^\delta}{\sigma}\right)^2. \tag{4.22}$$

By deriving with respect to x we obtain the following optimality condition

$$A^t Ax - A^t f^\delta = 0. \tag{4.23}$$

This is the normal equation for the least squares problem for the solution of $Ax = y^\delta$. Since A a is an ill conditioned matrix, the solution x of (4.23) might be heavily influenced by the noise.

As we saw in this example maximum likelihood estimation lead again to the normal equation, which might give a meaningless solution. Thus we turn our focus to Bayesian methods, in order to incorporate prior knowledge on the the desired solution.

Bayesian Methods

The central idea in Bayesian methods is to pick an estimator for the desired quantity given the measured data by using Bayes famous rule on conditional probabilities. Thomas Bayes' (1701-1761) version of the article *An Essay Towards Solving a Problem in the Doctrine of Chance* was edited by Richard Price and published posthumously in 1763 [14]. His famous result on conditional properties for events A and B is given by:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}. \tag{4.24}$$

For a given inverse problem

$$Ku = f \tag{4.25}$$

our goal is to use Bayes rule (4.24) to derive an estimator for u. One possibility is to maximize the a-posteriori probability

$$p(u|f). \tag{4.26}$$

By applying the Bayes rule, we can compute this probability as

$$p(u|f) = \frac{p(f|u)p(u)}{p(f)}. \tag{4.27}$$

For computing the maximum a posteriori (MAP) estimator we maximize this probability (4.27) now with respect to u; since $p(f)$ is a constant independent of u, this factor can be

neglected in the maximization process. Again we minimize the negative log-likelihood function, instead of maximizing (4.27). This leads to the problem

$$x_{\text{MAP}} := \min_u \left(-\log(p(f|u)) - \log(p(u)) \right). \tag{4.28}$$

While $p(f|u)$ is called the likelihood (compare the previous paragraph), $p(u)$ is called the prior probability. Dependent on the noise characteristics of the given problem, the negative log-likelihood is then given as a distance measure, e.g. the squared L^2 norm

$$-\log(p(f|u)) = \frac{1}{2\sigma} \int_\Omega (u-f)^2 \mathrm{d}x \tag{4.29}$$

or the Kullback-Leibler divergence

$$-\ln(p(f|u)) = \int_\Omega u - f \log(u) \mathrm{d}x. \tag{4.30}$$

The prior probability is often given by a Gibbs density [54] of the form

$$p(u) \propto \exp(-R(u)) \tag{4.31}$$

with R being a regularization term like total variation. We will explore different regularization terms, designed for specific tasks in imaging, in the forthcoming sections. If we choose e.g. a Gaussian Prior of the form

$$p(u) \propto \exp(-\|u\|^2) \tag{4.32}$$

and have data with Gaussian noise, then the Bayesian approach is strongly connected to the Tikhonov functional; compare [81, Chapter 3.4, Example 5].
Note that MAP-estimation is only one method in Bayesian modeling, there are others such as the conditional mean-estimate

$$x_{\text{CM}} = E\{x|f\} = \int_{\mathbb{R}^d} x p(x|f) \mathrm{d}x, \tag{4.33}$$

provided that the integral converges [81]. Since we focus on MAP-estimation methods in this thesis we will not elaborate further on this, but refer to [81] for more details. We have now stated the basic ideas of Bayesian modeling in inverse problems. See for example [81, 79, 140] for far more detailed introductions and discussion to statistical and especially Bayesian methods in inverse problems. Having described a framework, which allows us to incorporate noise characteristics as well as prior knowledge on the desired solution into the modeling of an ill-posed problem, we will now turn our focus on the two problems we are mainly concerned with throughout this thesis.

4.2 Reconstruction Problems

We start this section on reconstruction problems with formulating a general framework for our goal of reconstructing images from indirect measurements. In medical imaging reconstruction problems are for example given in:

- Computerized Tomography (CT)
- Emission Tomography (ET)
- Electric Impedance Tomography (EIT)

While in the first two modalities (a random sample of) the Radon transform is measured [40, 106], in EIT one tries to deduce the conductivity inside the body from the conductivity on the boundary [35].
We will now examine the building blocks of reconstruction problems:

- The output of a reconstruction method is the **reconstructed image**.
- Images are not measured directly, instead we have **indirect measurements** (for example on the boundary of the image domain).
- The **measurements** or **data** is the input in the reconstruction problem.

We will put these blocks into a mathematical framework. The indirect measurements mentioned above are recorded in the data space. Thus we have a forward operator

$$K : \mathcal{X} \to \mathcal{Y}, \tag{4.34}$$

mapping from a function space on the image domain $\mathcal{X}$ into the domain of measurements. While different imaging modalities might have different forward operators, usually the forward operator in different reconstruction problems is considered compact. Hence for given measurements f we have to solve the problem

$$Ku = f \tag{4.35}$$

for u. This is exactly the prototype of an inverse problem we introduced earlier. In this thesis we are mostly concerned with the Radon-Transform [106].

Definition 4.2.1 (Radon transform)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Schwartz function. For $\theta \in S^{n-1}$ and $s \in \mathbb{R}$ the Radon transform is given by

$$(Rf)(\theta, s) = \int\limits_{x\theta=s} f(x)\mathrm{d}x. \tag{4.36}$$

Remarks and Extensions.

- Schwartz functions are C^∞ functions with rapidly decreasing derivatives [106].

- The name of this transform is due to Johann Radon, who already gave a reconstruction formula in 1917 [118]. Being a completely technical problem then, the Radon transform found its application in the 1960s and 70s, when the first computer tomographs were designed by Cormack [38] and Hounsfield [78]. Both did not know about Radon's work and had to design reconstruction algorithms on their own.

The following example illustrates that direct inversion is not an option for compact operators.

Example 4.2.1 (Inversion of a compact operator)**.** Let $\mathcal{X}$ be an infinite-dimensional Hilbert space, $\mathrm{K} : \mathcal{X} \to \mathcal{X}$ a compact operator. For any orthonormal system $(o_n)_{n\in\mathbb{N}}$ we know that the sequence of images

$$(\mathrm{K}\, o_n)_{n\in\mathbb{N}}$$

is bounded, because K is compact. Therefore we find a subsequence n_j, such that

$$(\mathrm{K}\, o_{n_j})_{j\in\mathbb{N}}$$

converges, but in contrast, we have for any $k \neq l$:

$$\|o_k - o_l\|^2 = \langle o_k - o_l, o_k - o_l\rangle = \underbrace{\langle o_k, o_k\rangle}_{=1} - 2\underbrace{\langle o_k, o_l\rangle}_{=0} + \underbrace{\langle o_l, o_l\rangle}_{=1} = 2.$$

Thus the inverse K^{-1} cannot be continuous.

This example shows that image reconstruction is ill-posed in the sense of Hadamard (Definition 4.0.1). Since the presented modalities can not guarantee exact measurements instead of noisy versions of the projected image, we will adapt the methods from the previous section to the reconstruction problem.
Since our main application is emission tomography where a sample of the Radon transform is recorded, we focus on the Radon transform as forward operator.
In order to highlight possible difficulties in inverting the Radon transform, we study the inversion of a tomography matrix corresponding to a discrete version of the Radon transformation.

Example 4.2.2 (Discrete 2D-tomography)**.** We use the regularization toolbox by Per-Christian Hansen [71, 72] for an academic example. This package allows us to generate a 2D-tomography matrix for a 64×64 images and $64 \cdot 64 = 4096$ rays crossing the domain in random directions. This matrix is severly ill-conditioned, as we observe by a plot of the singular values:

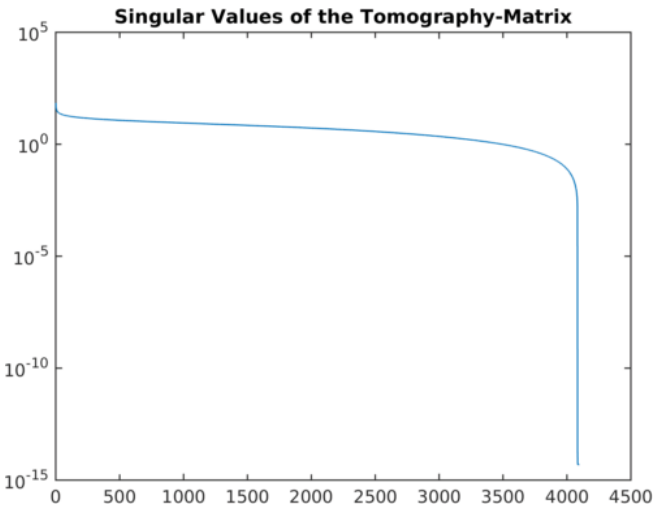
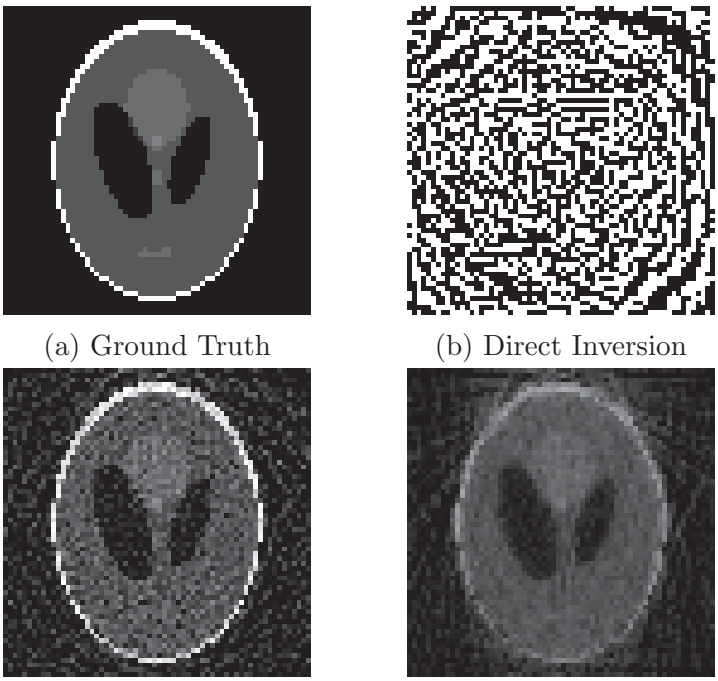


Figure 4.1: Logarithmic plot of all 4096 singular values of the tomography matrix. Note that the 500 smallest singular values decay rapidly towards 0. SVD computed by the regularization toolbox [71].

We aim to reconstruct the Shepp-Logan software phantom (intensities in the interval $[0, 1]$) provided by Matlab from measurements corrupted by Gaussian noise (with mean 0 and standard deviation $\sigma = 0.25$):



(a) Ground Truth

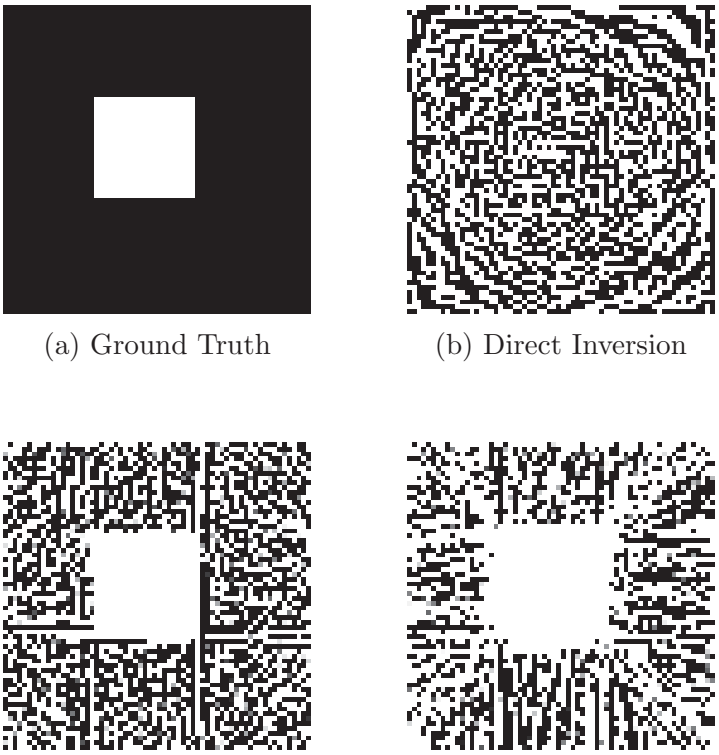
(b) Direct Inversion

(c) Truncated SVD ($J = 3500$)

(d) Tikhonov Regularization ($\alpha = 10$)

Figure 4.2: Reconstruction results for the Shepp-Logan-Phantom for different regularization methods. The direct inversion of the tomography matrix is meaningless.

There is no similarity between the ground truth and the solution obtained by direct inversion of the projection operator. In contrast truncated SVD and Tikhonov regularization allow us to see the structure of the object. Note that we did neither optimize the truncation index J nor the regularization parameter α. However the structure of the object is visible.
As a second example we aim to reconstruct an image consisting of two homogenous regions with intensities in the interval $[0, 10]$, which was again corrupted with Gaussian noise with mean 0 and standard deviation $\sigma = 0.25$:



(a) Ground Truth

(b) Direct Inversion

(c) Truncated SVD ($J = 3500$)

(d) Tikhonov Regularization ($\alpha = 10$)

Figure 4.3: Reconstruction of a single white square on black background with different regularization operators. Again the direct inversion leads to meaningless results.

As we see both regularization methods recover again the structure of the object, but we have a high amount of background noise.

The reconstruction example shows, that the regularization methods we presented earlier have difficulties in reconstructing homogeneous regions with sharp edges. In the next section we will present a regularization functional which will be able to yield sharp edges.

TV-Regularization for (Density) Images

As the reconstruction problem is stated exactly in the same way as the prototype of an inverse problem in the section before and we already applied these regularization methods, we do not discuss the presented regularization operators adapted to the reconstruction problem again. We rather focus on a specific regularization functional, which can be used as a norm penalty in the Tikhonov approach or for determining an a priori probability in a Bayesian framework.
In this section we want to present TV-regularization for density images, which will later correspond to the prior knowledge we impose on the image, because it favors homogeneous regions with sharp edges. We focus only on this regularization, since our application is in medical imaging and TV-regularization allows for discontinuous images with sharp edges. This regularization was first introduced for image denoising by Rudin, Osher and Fatemi in 1992 [127] and is now known as the ROF-model. Recently TV regularization has been applied to different tasks in imaging. We will now present the basics of the class of admissible functions for the method, namely the space of functions of bounded variation BV.
In fact there are many different equivalent ways to define $BV-$functions (compare [8, Chapter 10, Definition 10.1.1]), but we focus on the definition by Burger and Osher [26]. In order to define total variation, we consider the following spaces of test functions

$$\tilde{Z} := \{\nabla \cdot g \,|\, g \in C_0^\infty(\Omega; \mathbb{R}^d, \|g\|_\infty \leq 1\}, \tag{4.37}$$

where $\|g\|_\infty$ is given by

$$\|g\|_\infty := \operatorname*{ess\,sup}_{x\in\Omega} \sqrt{\sum_{i=1}^{n} g_i(x)^2}. \tag{4.38}$$

The BV seminorm can now be defined with help of these test functions.

Definition 4.2.2 (BV seminorm and functions of bounded variation)**.** Let

$$u : \Omega \subset \mathbb{R}^d \to \mathbb{R}.$$

The BV-seminorm is then given by:

$$|u|_{BV(\Omega)} := \sup_{z\in\tilde{Z}} \int_\Omega uz \mathrm{d}x. \tag{4.39}$$

Consequently we define the space of functions with bounded variation $BV(\Omega)$ by

$$BV(\Omega) := \{u \in L^1(\Omega) \,|\, |u|_{BV(\Omega)} < \infty\} \tag{4.40}$$

This space is equipped with the following norm:

$$\|u\|_{BV} := \|u\|_1 + |u|_{BV(\Omega)} \tag{4.41}$$

Remarks and Extensions. The BV-seminorm is lower semicontinuous in the L^1_{loc} topology [5, Remark 3.5].
If $u \in W^{1,1}$ then the BV-seminorm is given by:

$$|u|_{BV(\Omega)} = \int_\Omega |\nabla u| \mathrm{d}x. \tag{4.42}$$

The definition above is the isotropic version of the TV-regularization. By replacing the Euclidean (4.38) norm of the gradient with an appropriate functional anisotropic versions of the TV regularization can be derived [26]. Although leading to the same space $BV(\Omega)$, the choice of the norm can have significant impact on the reconstruction - compare [26, Figure 4] and the references given therein.

Properties of BV-functions

In the following we collect some useful properties from [25] and [8]. One of our basical tools in the analysis of functionals is the direct method in the calculus of variations. For using this method we have to verify lower semicontinuity and compactness of sublevel sets. The latter one can be shown by the Banach-Alaoglu theorem (compare Theorem 3.4.7), if the functional is defined on the dual of a Banach space. We present shortly the properties of BV functions which are crucial to the analysis we carry out in Chapter 7. For the analysis in BV one often decomposes a function $u \in BV(\Omega)$, such that

$$u = u_0 + u_1 \tag{4.43}$$

with

$$\int_\Omega u_0 \mathrm{d}x = 0 \qquad\qquad |u_1|_{BV(\Omega)} = 0. \tag{4.44}$$

Remarks and Extensions. This decomposition can be achieved for any function $u \in BV(\Omega)$ by setting

$$u_0 := u - \int_\Omega u \mathrm{d}x \qquad\qquad u_1 \equiv \int_\Omega u \mathrm{d}x. \tag{4.45}$$

In doing so much of the analysis can be carried out in the space

$$BV_0(\Omega) := \{u \in BV(\Omega) | \int_\Omega u \mathrm{d}x = 0\}. \tag{4.46}$$

Proposition 4.2.3. *For $BV_0(\Omega)$ we can construct a space Z, such that $BV_0 = Z^*$. For the space*

$$Z_0 := \{\nabla \cdot g | g \in C_0^\infty(\Omega, \mathbb{R}^d)\} \tag{4.47}$$

equipped with the norm

$$\|p\|_Z := \inf_{\substack{g \in C_0^\infty(\Omega, \mathbb{R}^d) \\ \nabla \cdot g = p}} \|g\|_\infty, \tag{4.48}$$

we denote the completion by $Z := \overline{Z_0}$. *Then* BV_0 *is the dual of this space, namely:*

$$BV(\Omega)_0 = Z^*. \tag{4.49}$$

Proof. See [26, Proposition 3.4]. □

Note that also $BV(\Omega)$ is the dual of a separable space, see for example [5, Remark 3.12]. According to [5] we can define a weak-star convergence, which is consistent with BV being the dual of a Banach space:

Definition 4.2.4 (Weak-star convergence in BV)**.** Let $u, u_n \in BV(\Omega)$. We say

$$u_n \rightharpoonup^* u$$

if the function converges strongly

$$\lim_{n\to\infty} \|u_n - u\|_1 = 0 \tag{4.50}$$

as well as the differential weakly:

$$\lim_{n\to\infty} \int_\Omega \varphi(Du_n - Du)\mathrm{d}x = 0 \qquad \forall \varphi \in C_0(\Omega). \tag{4.51}$$

Note that even the weak star convergence, which is implied by strong and even weak convergence, guarantees strong convergence in L^1. As we will see later, this will be useful to prove convergence properties of compositions of functions. Furthermore we can approximate any BV-function by a sequence of smooth functions:

Theorem 4.2.5 (Approximation by smooth functions)**.** *Let* $u \in L^1(\Omega; \mathbb{R}^m)$ *Then*

$$u \in BV(\Omega, \mathbb{R}^m),$$

if and only if there exists a sequence u_n *in* $C^\infty(\Omega, \mathbb{R}^m)$ *converging to* u *in* $L^1(\Omega, \mathbb{R}^m)$ *and satisfying*

$$L := \lim_{n\to\infty} \int_\Omega |\nabla u_n| \mathrm{d}x < \infty \tag{4.52}$$

Proof. [5, Thm. 3.9]. □

With TV-regularization we have described a regularization functional, which favors sharp edges and homogeneous regions, which is favorable in medical imaging. In order to cope with regions where the intensity changes smoothly between the edges, the ideas presented above can be generalized to higher derivatives [19]. Additionally classical total variation regularization can produce an unwanted staircasing effect, which can be removed by adding higher order terms [34].

4.3 Image Registration

Image registration is the task of aligning two or more images, which are mismatched because of a different time of taking, view point or even image modality. We will only briefly present the basics of image registration, for a detailed overview see for example [100] and the references therein. We start by giving the image registration problem in Modersitzki's formulation [101]:

Image Registration Problem

Find a **reasonable transformation**, such that a transformed version of a **template image** is **similar** to a **reference image**.

We will now define a functional, such that the minimizer of the functional solves the registration problem. In this problem several words are in bold print; these will correspond somehow to the building blocks of the functional:

- **Template** f_0 and **reference** image f_1 are the images to be registered.
- The **transformation** y shall be found as the minimizer of the functional (4.53).
- The **similarity** is measured with a distance measure D.
- The **reasonability** of the transformation corresponds to the regularization energy of the transformation $S(y)$, such that we call a transformation reasonable, if the regularization energy is sufficiently small.

Building these together we can define the following cost functional:

Registration Functional

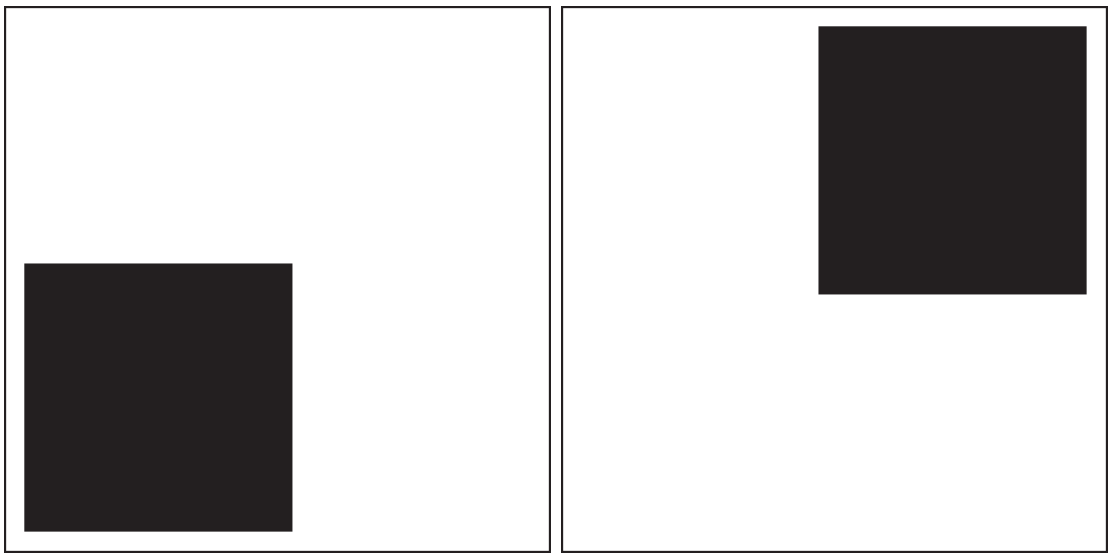
$$J(y) := D(f_0(y), f_1) + S(y) \tag{4.53}$$

Note that template and reference image are often dentoted by T and R, but since we will use the letter T for for transformation operators later, we do not adopt this notation. We start by motivating the use of regularization energies and continue with presenting all building blocks of the registration functional shortly.

4.3.1 Image Registration: An Ill-Posed Problem

To give a more visual impression of the need for special treatment of image registration we present another example by [48]:

Example 4.3.1 (Registration of two squares)**.** The following example illustrates the ill-posedness of image registration:



(a) Reference (b) Template

Figure 4.4: A simple registration problem: The template image shall be aligned with the reference image.

As we see directly, there are several ways to align both images: A rotation of the whole image as well as a translation leads to an exact registration result. In fact any transformation that maps the points from the upper right square to the lower left is exact. Because this latter solution is in most cases not desired, regularization has to be employed. Note that these registration possibilities may also cause several problems in the evaluation of an estimated motion vector field.

As we motivated with this short example, in most cases regularization becomes inevitable for image registration. This regularization can be linked to a priori information on the desired quantity in the Bayesian modeling. In the following we will outline the central building blocks of the registration functional.

4.3.2 Transformation Models

There are two transformation models mainly used in image registration. The first one is the composition of an image f_0 with the transformation y

$$T_y^{ic} f_0 := f_0(y). \tag{4.54}$$

The composition is called *intensity-constant* transformation operator, because the intensity of the transformed image is not modulated.
While doing registration on density images one might desire, that the integral of the transformed image is the same as the integral of the non-transformed. Motivated by the transformation theorem this leads to the so called *mass-preserving* transformation, defined for transformations with positive Jacobian determinants [128]

$$T_y^{mp} f_0 := f_0(y) \det(\nabla y). \tag{4.55}$$

Note that each transformation operator is linear with respect to the image it is applied to for any fixed transformation. We will use this property later for deriving a registration framework with noise modeling.

4.3.3 Distance Measures

In this section we will briefly outline some distance measures used in image registration, see e.g. [101] for more details. The first and most intuitive way to define a distance measure is to compare the intensity of two images pixelwise:

Definition 4.3.1 (Sum-of-squared-differences)**.** For two images $f_0, f_1 \in L^2\Omega$ the sum-of-squared-differences distance measure (SSD-measure) is defined by

$$D^{SSD}(f_0, f_1) := \int_\Omega (f_0 - f_1)^2 \mathrm{d}x. \tag{4.56}$$

Naturally this distance measure is best suited for two images from the same modality, but registering two images from different modalities is a common task in image registration. Thus there are several approaches to design distance measures for multi-modal image registration, such as [124, 66]. In this case the central idea in the latter is that even if intensities might not match, the edges of two aligned images from different modalities are still in the same place. Detecting edges via the image gradient leads to the following distance measure:

Definition 4.3.2 (Normalized Gradient Field)**.** For two images$f_0, f_1 \in W^{1,2}(\Omega)$ the normalized gradient field distance measure (NGF-measure) is defined by:

$$D^{NGF} := \int_\Omega 1 - \left(n(f_0)^t n(f_1)\right)^2 \mathrm{d}x \tag{4.57}$$

with

$$n(T) := n(f_0, \eta) := \frac{\nabla f_0}{\sqrt{|\nabla f_0|^2 + \eta^2}}. \tag{4.58}$$

The parameter η is an edge parameter, which determines what is considered to be an edge and what is considered to be within the noise level.
We restrict the presentation of distance measures to these two examples. There is a variety of distance measures used in the literature. See for example [101] for further measures like mutual information or normalized cross correlation.
Having presented the basics of image registration, we will use this as a tool in the later parts of the thesis. A possibility how to actually solve the image registration problem will be presented in the Chapter 5 dealing with the numerical basics.

4.3.4 Parametric and Non-Parametric Image Registration

We mainly distinguish between two types of image registration: parametric and non-parametric [101]. While non-parametric transformations are completely arbitrary parametric transformations can by described by few parameters:

Definition 4.3.3 (Parametric-transformations)**.** Let $(\phi_l)_{l=1,..,m}$, be a finite set of basis functions. A transformation $y : \mathbb{R}^d \to \mathbb{R}^d$ is called parametric (with respect to the (ϕ_l)) iff y can be expressed as a linear combination by parameters ω_l, such that

$$y(x) = \sum_{l=1}^{m} \omega_l \phi_l(x). \tag{4.59}$$

We will enlighten this definition by the following example presenting one of the most common parametric registration methods:

Example 4.3.2 (Affine linear transformations in 2D)**.** Let $A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \in \mathbb{R}^{2\times 2}$ be a regular matrix with $a_{i,j} \neq 0$. Furthermore let $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in \mathbb{R}^2$ with $b_i \neq 0$ a translation vector. Then the affine linear transformation y, induced by A and b is given by:

$$y : \mathbb{R}^2 \to \mathbb{R}^2 \qquad x \to Ax + b. \tag{4.60}$$

We define the following basis functions:

$$\begin{aligned} \phi_1(x) &= \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \qquad \phi_2(x) = \begin{pmatrix} x_2 \\ 0 \end{pmatrix} \qquad \phi_3(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \phi_4(x) &= \begin{pmatrix} 0 \\ x_1 \end{pmatrix} \qquad \phi_5(x) = \begin{pmatrix} 0 \\ x_2 \end{pmatrix} \qquad \phi_6(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned} \tag{4.61}$$

Thus the transformation is given by:

$$y(x) = a_{1,1}\phi_1(x) + a_{1,2} + \phi_2(x) + b_1\phi_3(x) + a_{2,1}\phi_4(x) + a_{2,2}\phi_5(x) + b_2\phi_6(x) \tag{4.62}$$

The linear function can be described with six basis functions and parameters. Note that we made the assumption on regularity and non-zero components of the vector only to guarantee that we need in fact all basis vectors to parametrize the transformation.

In registration pipelines the parametric estimation is often used as a pre-registration step to obtain the global behavior of the transformation, since this can be done with low computational costs. Subsequently, a non-parametric registration is usually performed in order to explore the local behavior of the desired transformation.

4.3.5 Regularization for Transformations

Since regularization is already a central component of the registration problem formulation by Modersitzki we presented at the start of this section, we will not carry out a Bayesian approach in order to deduce the framework we present here. However we will do this in a generalized form in Chapter 6. In this paragraph we focus on some common regularizers used in image registration.
Many of the regularization energies used in image registration are derived from physical properties of the transformation. See for example [100] for a detailed modeling and an elaborate explanation of the underlying physical basics. In order to keep this paragraph short, we focus only on elastic and hyperelastic regularization. See [100] for a thorough illustration of curvature and diffusive regularization.
The central idea of elastic regularization is, that the regularization energy of a transformation is given by the elastic potential of the transformation. In his two books on image registration Modersitzki presents two different versions of the elastic potential with positive Lamé constants μ and λ. In his first book [100] he uses the following version

$$S_1^{elas}(y) = \int_\Omega \frac{\mu}{4} \sum_{j,k=1}^{d} (\partial_{x_j} y_k(x) + \partial_{x_k} y_j(x))^2 + \frac{\lambda}{2} (\nabla \cdot y(x))^2 dx, \tag{4.63}$$

while in his second book [101] he proposes

$$S_2^{elas}(y) = \int_\Omega \mu \sum_{e=1}^{d} \langle \nabla y_e, \nabla y_e \rangle + (\mu + \lambda)(\nabla \cdot y)^2 dx. \tag{4.64}$$

Note that sometimes instead of the transformation y the displacement

$$y(x) - x \tag{4.65}$$

is regularized. In doing this one can achieve that the identity (not transforming the image) has zero regularization energy. While both versions of the elastic potential look similar, the first one differs from the second by the occurence of mixed terms $\partial_{x_j} y_k \partial_{k_k} y_j$. This has direct consequences for verifying coercivity properties of the regularizer: While coercivity can be directly deduced for the latter, coercivity is a consequence of Korn's inequality [36, Thm. 6.3-3] for the first. Nevertheless both versions of the elastic potential behave similar as a regularizer.
These somehow classic regularizers may cause severe problems, if the transformation is expected to be (at least locally) invertible. Especially elastic regularization can lead to non-invertible transformations [25]. Therefore we will focus on regularization functionals, which yield finite energy only for locally invertible transformations.

Hyperelastic Regularization

The central idea of hyperelastic regularization is to consider energy functionals of the form

$$S(y) = \int_\Omega g(\nabla y, \mathrm{cof}(\nabla y), \det(\nabla y))\mathrm{d}x. \tag{4.66}$$

The function g is chosen to guarantee certain features for the class of admissible transformations. The hyperelastic energy we use throughout this thesis is given by

$$S^{\mathrm{hyper}}(y) \;=\; \int_\Omega \alpha_1 \,\mathrm{len}(\nabla y) \;+\; \alpha_2 \,\mathrm{surf}(\mathrm{cof}(\nabla y)) \;+\; \alpha_3 \,\mathrm{vol}(\det(\nabla y))\,\mathrm{d}x\ , \tag{4.67}$$

with the penalty functions

$$\mathrm{len}(s) = \|s - \mathcal{I}\|^2_{\mathrm{Fro}}\ , \qquad \mathrm{surf}(s) = \left(\|s\|^2_{\mathrm{Fro}} - 3\right)^2, \qquad \mathrm{vol}(s) = \frac{(s-1)^4}{s^2}\ .$$

The three terms punish deviations from the identity related to length, surface and volume; see [25] for more details. This regularization energy directly implies a certain set of admissible transformations. Together with a detailed introduction of the hyperelastic regularization energy Ruthotto [129] gave the following set of admissible transformations:

$$y_k \in \mathscr{A}$$

with

$$\mathscr{A} := \{y \in \mathscr{A}_0 : |\int_\Omega y(x)dx| \leq \mathrm{vol}(\Omega)(M + \mathrm{diam}(\Omega))\} \tag{4.68}$$

where Ω is bounded by M, and $\mathscr{A}_0$ is defined by:

$$\begin{aligned}\mathscr{A}_0 := \{y \in W^{1,2} : &\mathrm{cof}(\nabla y) \in L^4(\Omega, \mathbb{R}^{3\times 3}) \\ &\det(\nabla y) \in L^2(\Omega, \mathbb{R}), \det(\nabla y) > 0 \text{ a.e.}\}\end{aligned}$$

Remarks and Extensions. If $y \in \mathscr{A}_0$ we call y an **admissible transformation**. Note that all admissible transformations fulfill the conditions on the cofactor in Proposition 3.2.9 and Theorem 3.3.8.

As we see the admissible transformations have strict positive Jacobians a.e., which is sufficient for the transformation to be locally invertible. Note that this deduction is not trivial: Since by the Sobolev embedding theorem (Theorem 3.1.11) an admissible function does not need to be continuous we cannot use the implicit function theorem to deduce the existence of a local inversion. Even worse, the standard theory about (local) invertibility of Sobolev mappings is focussed on mappings in $W^{1,d}$ (see for example [49, 87]), so this theory would only be applicable for $d = 2$. However, we can use some results from the theory of Cartesian currents we mentioned briefly earlier and use the fact, that:

$$\mathscr{A} \subset \mathscr{A}_{d-1,d-1}(\Omega, \mathbb{R}^m) \tag{4.69}$$

with $\mathscr{A}_{d-1,d-1}(\Omega, \mathbb{R}^m)$ as being defined before. Then a result from Müller [99] yields the closedness of the graph [59] of the transformation y and thus y is a weak local diffeomorphism as defined by Giaquinta et al. [61]. We will not elaborate on this further, because we used this argumentation only to demonstrate that we can expect to have local invertibility and that this property is not directly guaranteed by the positivity of the Jacobian determinant.
We conclude this short course on hyperelastic regularization by stating the convergence properties, Ruthotto proved [129]:

Theorem 4.3.4 (Convergence properties of admissible transformations)**.** *Let Ω be a domain with a C^1 boundary and y_k, y be admissible transformations. Then we have:*

$$\left.\begin{array}{ll} y_k \rightharpoonup y & \text{in } W^{1,2}(\Omega, \mathbb{R}^3) \\ \operatorname{cof}(\nabla y_k) \rightharpoonup H & \text{in } L^4(\Omega, \mathbb{R}^{3\times 3}) \\ \det(\nabla y_k) \rightharpoonup v & \text{in } L^2(\Omega, \mathbb{R}) \end{array}\right\} \Rightarrow \begin{cases} H = \operatorname{cof}(\nabla y) \\ v = \det(\nabla y) \end{cases} . \tag{4.70}$$

Proof. See [129, Chapter 3, Theorem. 4]. □

5 Numerical Methods to Minimize Functionals

This chapter deals with numerical methods to solve the optimization problems we introduced in the previous chapter. In general there are two different ways to obtain solutions for these problems: The First-Discretize-then-Optimize approach, where the problem is first discretized and this version is optimized afterwards. In contrast to this the First-Optimize-then Discretize approach is characterized by determining optimality conditions for the continuous problem followed by solving these conditions via discretization techniques.
In the following we will illustrate the two different ways and outline some realizations. Since the focus of this work lies on the First-Discretize-then-Optimize method, it will be described in more detail.

5.1 Discretize-then-Optimize Methods

The central idea of the First-Discretize-then-Optimize approach is to derive a sequence of discretized versions of the functional to be minimized. Each discretized version is then solved with the help of tools from optimization. We start by deriving a discretized version of the functional to be minimized.

5.1.1 Discretization

As we see for example in the registration functional (4.53) many functionals in image processing contain integrals. For the discretization we divide the image domain into small disjoint boxes B_i for $i = 1, ..., N$, such that

$$\bigcup_i B_i = \Omega \qquad B_i \cap B_j = \emptyset \quad \forall i \neq j. \tag{5.1}$$

These boxes are called pixels, or in the case of three dimensional domains sometimes voxels. With this discretization of the image domain the discretization of a continuous image is straightforward: For a continuous image ρ we set the value in b_i in the i-th pixel as

$$b_i := \int_{B_i} \rho(x) \mathrm{d}x. \tag{5.2}$$

Taking this definition of the discrete image into account, we continue by introducing discrete versions of transformations in the next paragraph.

Grids

A discrete version of a transformation is given by the values on a point sample of the domain, the so-called grid. We only briefly mention the most important notions and construction methods for grids; see for example [65, 101] for a description in more detail. In the following we distinguish between the following three types of grids:

- cell-centered grids
- nodal grids
- staggered grids

We will explain these basic grids in two dimensions, since the extension to higher dimensions is straightforward. As a preparation we start with grids in one dimension: Let $\Omega = [0, \omega_1]$ and n the number of cells Ω is divided into. Then the cell centered grid is given by

$$x_{cc}(i) := \left(i - \frac{1}{2}\right) h \qquad i = 1, ...n \qquad \text{with cell size} \qquad h = \frac{\omega_1}{n}. \tag{5.3}$$

Note that the grid points of the cell-centered grid, are the center points of the grid cells. In contrast the grid points of the nodal grid are the nodes of the grid cells

$$x_n(i) := ih \qquad i = 0, ...n \qquad \text{with cell size} \qquad h = \frac{\omega_1}{n}. \tag{5.4}$$

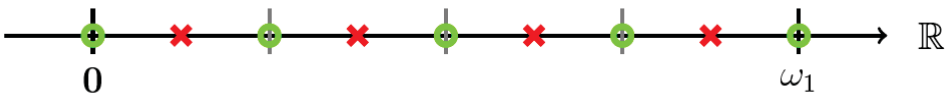


Figure 5.1: Nodal (red crosses) and cell-centered (green circles) grid in one dimension. Graphic from [113].

Having defined these grids in one dimension, the extension to higher dimensions can be done in a natural way: We again assume a rectangular domain

$$\Omega := [0, \omega_1] \times [0, \omega_2].$$

Here we can define the cell-centered and nodal grid for each component of the Cartesian product. Cell-centered and nodal grids in two dimensions are then obtained via the Cartesian product of the one-dimensional ones. If we combine a cell-centered and a nodal grid, then we get the so called staggered grids.

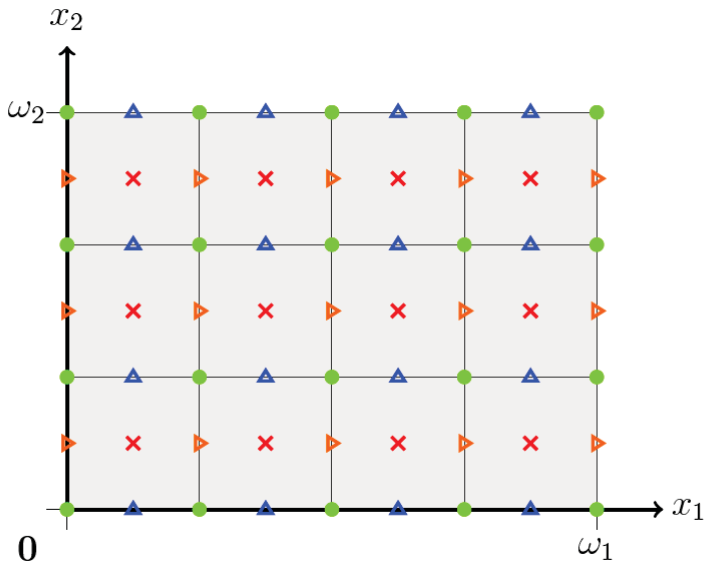


Figure 5.2: Nodal (red crosses), cell centered (green circles) and staggered (triangles) grids in two dimensions. Graphic from [113]

Note that all these grids are used in our registration framework, so we need a grid to grid interpolator, which changes between the different grid types. We will not focus on this interpolator, but rather use it as a tool; see [113, 101] for further details on this matter. Having explained these basic grids, we now discuss numerical integration and differentiation.

Numerical Integration and Differentiation

As we have seen in the previous chapter, many of the distance measures in image registration are given as an integral over an integrand function. The regularizers also involve integration of the norm of some derivatives of the transformation. Therefore we start by shortly outlining numerical integration and differentiation.
For a sufficiently regular function u given on a cell centered grid $x_{cc}(i)$ with cell size h_{cc} we perform numerical integration with a simple midpoint quadrature rule (compare for example [42])

$$\int_\Omega u(x)\mathrm{d}x = \sum_i h_{cc} u(x_{cc}(i)) + \mathcal{O}(h_{cc}). \tag{5.5}$$

Note that there are several more accurate quadrature formulas, but for the ease of computations we stick to this simple midpoint quadrature formula. See for example [42] for a detailed presentation of various quadrature formulas.
Having shown a simple but yet useful way for numerical integration, we turn our focus towards numerical differentiation. We only give a short overview; detailed descriptions can be found in [24]. Again we choose an intuitive way to approximate the derivative of

a function: Inspired by the definition of differentiability we use the difference quotient Δ for the approximation of the derivative of a function u for $h > 0$

$$\Delta_f u(x) := \frac{u(x+h) - u(x)}{h} \qquad \Delta_b u(x) := \frac{u(x) - u(x-h)}{h}. \tag{5.6}$$

While these so called forward and backward difference quotients have an error in $\mathcal{O}(h)$, we can define the central difference quotient with an error in $\mathcal{O}(h^2)$ by

$$\Delta_c u(x) := \frac{u(x+h) - u(x-h)}{2h}. \tag{5.7}$$

Note that this property of the central difference quotient is one of the reasons, we choose different transformation grids for some regularizers: If a transformation is given on a nodal grid, we are able to compute an approximation for the derivative on a cell centered grid with error in $\mathcal{O}(h^2)$ by using the central difference quotient. See for example [24] for further details on numerical differentiation and [101] for a detailed description of the relation between regularization in image registration and the choice of the grid type, representing the discrete transformation.

5.1.2 Line Search and Trust Region Methods

By applying the described discretization to a functional to be minimized (e.g. the registration functional) we end up with a finite, but maybe high-dimensional optimization problem.
Let us assume we want to register two three-dimensional images with resolution $175 \times 175 \times 47$ voxels, then the desired (cell-centered) grid has roughly four million ($175 \cdot 175 \cdot 47 \cdot 3 = 4318125$) unknowns. Since analytical minimization is not an option, we need methods to deal with such big optimization problems. In this section we will briefly outline the basics of two widely used classes of optimization strategies: line search and trust region methods. We will roughly follow the presentation from [110]; see also there for a far more detailed course on numerical optimization.
While line search methods find a search direction (for example $s = -\nabla J(x)$) and try to minimize the effective function along this line, in trust region methods the function is described via a model function near the actual point and the minimizer of the model function is the next iterate.
The trust region minimization follows roughly the following scheme:

1. Choose an initial point x_0 and trust region size Δ_0.
2. Define the model function at the current point.
3. Minimize the model function inside the trust region.
4. Adjust the size of the trust region Δ_k and take the minimizer as next iterate, if the step is accepted. Go to step 2.

The key questions here are:

- How do we choose the model function?
- How do we solve the restricted optimization problem in step 2?
- How is the size of the trust region adjusted?
- When is a step accepted?

We will not answer these questions in detail, but we will shortly present the ideas how these questions can be answered. For a detailed description and comparison between different possibilities how to answer the questions see [110].
The model function m_k for a function j at the iterate x_k is usually chosen as a quadratic approximation

$$m_k(x_k + s) := j(x_k) + s^T \nabla j(x_k) + \frac{1}{2} s^T B_k s, \tag{5.8}$$

where the matrix B_k is either the Hessian $\nabla^2 J(x_k)$ or an appropriate approximation.
Solving the trust region subproblem is the most complex step in the algorithm. Since we have to deal with a restricted optimization problem, the minimization method has to be chosen carefully. Since we will not use trust region methods in our optimization scheme, we only present the Cauchy point (approximate) solution to the problem. For that we only solve

$$s_k := \arg\min_s j(x_k) + \nabla j(x_k)^T s \qquad \text{s.t.} \quad \|s\| \leq \Delta_k. \tag{5.9}$$

Here the solution is directly given by

$$s_k := -\frac{\Delta_k}{\|\nabla j(x_k)\|} \nabla j(x_k). \tag{5.10}$$

Now we calculate the scalar τ_k that minimizes $m_k(\tau s_k)$ and choose the Cauchy step as

$$s_k^C := \tau_k s_k. \tag{5.11}$$

Note that the Cauchy step only solves a restricted linear problem instead of the quadratic model. This can be improved via the Dogleg- or Steihaugs CG-method [110].
The last two questions can be answered together: For a step s_k we can define the ratio

$$r_k := \frac{j(x_k) - j(x_k + s_k)}{m_k(0) - m_k(s_k)} \tag{5.12}$$

between actual reduction (numerator) and predicted reduction (denominator). Since the denominator is always non-negative, a negative ratio always indicates an (unwanted) increase of the function. Thus a step is accepted, if the ratio is greater than some specified value. If this ratio is in addition close to 1, the model fits well to the function and the trust region can be enlarged - in contrast the size of the trust region is decreased, if the ratio is negative or close to zero.

In contrast in line search methods we are looking for a step length $\hat{\alpha}$, such that for a given search direction s:

$$\hat{\alpha} := \arg\min_{\alpha} j(x_k + \alpha s) \tag{5.13}$$

If this equation is solved exactly in an algorithm, this is called an exact line search method. Since it is mostly too time consuming to solve (5.13) exactly, there are several conditions employed, which can guarantee sufficient descent in inexact line search methods. Now we discuss some of these conditions before we present some choices of descent directions:

Strategy	**Condition(s)**	**Parameter choice**
Armijo	$j(x+\alpha s) \leq +\sigma\alpha(\nabla j(x)^T s)$	$\alpha, \sigma \in (0,1)$
Goldstein	$j(x) + (1-\sigma)\alpha\nabla j(x)^T s \leq j(x+\alpha s)$ $j(x+\alpha s) \leq j(x) + \sigma\alpha\nabla j(x)^T s$	$\sigma \in (0, 0.5)$
Wolfe-Powell	$j(x+\alpha s) \leq \sigma\alpha\nabla j(x)^T s$ $\nabla j(x+\alpha s)^T s \geq \rho\nabla j(x)^T s$	$\sigma \in (0, 0.5)$ $\rho \in [\sigma, 1]$

Table 5.1: Some common step length strategies. See for example [110].

Note that global convergence properties of line search methods rely on Zoutendijk's theorem; see for example [110, Chapter 3.2] for further details, since a discussion about convergence properties goes beyond the scope of this chapter. Note that the Goldstein conditions are not well suited for quasi-Newton methods [110]. Furthermore finding step length parameter fulfilling the Wolf conditions usually requires more computational effort then finding Armijo step lenghts by backtracking line search. Thus we will use the Armijo condition in our quasi-Newton optimization framework.
Having presented the most common step length strategies we will now turn our focus on the determination of an appropriate search direction. Inspired by Newton's method, there is a class of search directions we will describe in more detail in the next section:

Newton Type Methods

The central idea in Gauss-Newton type methods is, to use an approximation for the function to be minimized and use this approximation to compute a descent direction. For example for a function $j \in C^3$ we can do the second order Taylor approximation at a point x

$$j(x+s) = j(x) + \nabla j(x)s + \frac{1}{2}s^T \nabla^2 j(x)s + \mathcal{O}(s^3). \tag{5.14}$$

The central idea is then to minimize this approximation with respect to s. Since this is a quadratic form, it is known, that (for $\nabla^2 j$ positive definite) the minimizer $\hat{s}$ is given by:

$$\hat{s} = -\nabla^2 j(x)^{-1} \nabla j(x). \tag{5.15}$$

If we want to determine the exact Newton search direction we face a problem: Since we are mostly dealing with big problems it can get very complicated to invert the Hessian or even to store it. In fact often a conjugate gradient approach [98] is used to invert the Hessian. If we take a look at the PET-registration, which is an application for the motion-corrected reconstruction framework we present later, the Hessian in this case would be of the size $4,318,125 \times 4,318,125$, with $18,646,203,515,625$ elements in total. Assume that each entry is saved as a double with 8 bytes, this requires $149,169,628,125,000$ bytes, respectively 149169628.125 gigabytes. Even though the amount of memory needed can be dramatically decreased by using the symmetry and possible sparsity of the Hessian, storing the full Hessian is often not an option. Because of that quasi-Newton methods, where the Hessian matrix is approximated, are often used.
We will now present the Broyden-Fletcher-Goldfarb-Shanno method (BFGS), which is the most popular quasi-Newton-Algorithm and will be used in our minimization framework. For an initial estimation of the Hessian B_0, we update the approximation of the Hessian by using only information from the gradient

$$B_{k+1} := B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\gamma_k^T \gamma_k}{\gamma_k^T s_k}, \tag{5.16}$$

with

$$s_k := x_{k+1} - x_k \qquad \gamma_k := \nabla j(x_{k+1}) - \nabla j(x_k). \tag{5.17}$$

Note that B_{k+1} stays symmetric positive definite for a convex function, if $\gamma_k^T s_k > 0$. Unfortunately, this is not the case if we deal with a non-convex objective function. To overcome these difficulties a modified BFGS-approach has been proposed [90]. The key idea is to modify the γ_k slightly to

$$\tilde{\gamma}_k := \gamma_k + \left(1 + \max\left\{-\frac{\gamma_k^T s_k}{\|s_k\|^2}, 0\right\}\right) \|\nabla J(x_k)\| s_k \tag{5.18}$$

and then use (5.16) with $\tilde{\gamma}_k$ instead of γ_k. According to the authors this modified method also has a global convergence property if the function is not necessarily convex, but has only Lipschitz-continuous gradients. [90, Theorem 3.4]. By applying this method a superlinear convergence rate may be achieved. Note that this convergence property is only given, when the original BFGS-method converges.
Another useful property of the BFGS update is, that with the help of the Sherman-Woodbury-Morrison formula [110, Appendix 2] the inverse can directly be obtained

from the previous iterate. Let B_k^{-1} be the inverse for B_k and B_{k+1} be constructed by (5.16), then

$$B_{k+1}^{-1} = \left(I - \frac{s_k \gamma_k^T}{\gamma_k^T s_k}\right) B_k^{-1} \left(I - \frac{\gamma_k s_k^T}{\gamma_k^T s_k}\right) + \frac{s_k s_k^T}{\gamma_k^T s_k}. \tag{5.19}$$

Another benefit of the BFGS-method is, that one only has to store the gradients and iterates of the previous steps, instead of the full Hessian. But even this can require to a huge amount of storage capacities. Inspired by some limited memory CG-algorithms [114, 136] Nocedal presented a limited memory version of the BFGS algorithm [109]. The key idea is, to store only a certain number of vectors generating the BFGS matrix and just delete the oldest, once the maximum capacity is reached. The l-BFGS method converges, if certain convexity assumptions are fulfilled; see [92, Theorem 6.1].

5.1.3 Multilevel Methods

Another concept for solving the registration problem are multilevel methods. Inspired by multigrid strategies for partial differential equations (see [67, 141] for an introduction) the idea is to solve the problem on a coarse grid with less computational cost and use this solution as a initial guess to solve the problem on a finer grid.

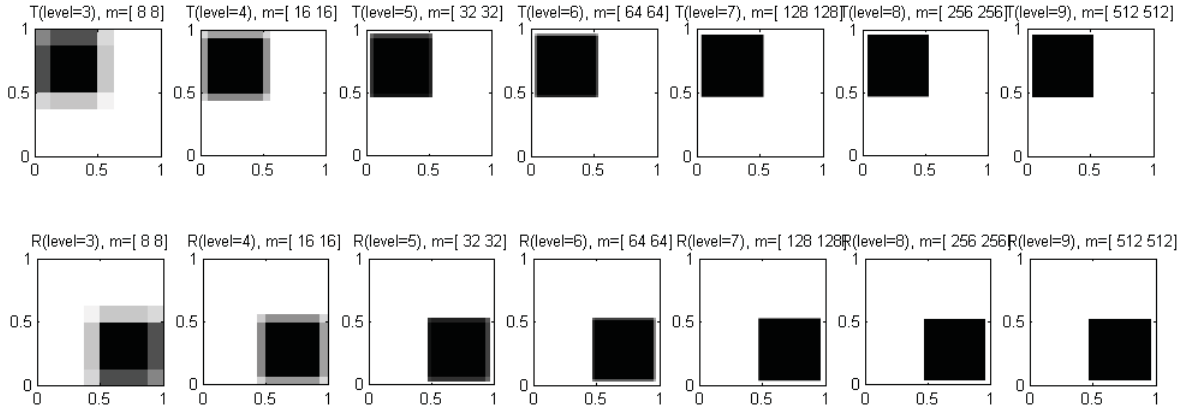


Figure 5.3: Multilevel representation of the registration problem from Figure 4.4. Note that the squares overlap at the lowest level (left), while they are disjoint at the highest (right). Plots by the FAIR toolbox [101]

This approach has the advantage, that it decreases the probabilty of finding local instead of global minima. While the squares (Figure 5.3) on the highest level are completely disjoint, the identity is a local minimum: Moving the square a bit does not improve the data fidelity, but results in a higher regularization energy. In contrast the overlap of the squares on the lowest level can push the template image in the right direction.

Another aspect of the multilevel approach is the lessening of noise impact, as we motivate in the following example:

Example 5.1.1 (Multilevel with Gaussian noise in 2D)**.** Let $A = a_{i,j}$ be a two dimensional image corrupted with independent identically distributed Gaussian noise, such that

$$a_{i,j} = \hat{a}_{i,j} + \epsilon_{i,j}. \tag{5.20}$$

Here $\hat{a}$ denotes the uncorrupted image and ϵ is standard normally distributed with mean zero and variance σ^2. The image $B = b_{i,j}$ at the level below is given by

$$b_{i,j} = \frac{a_{k,l} + a_{k+1,l} + a_{k,l+1} + a_{k+1,l+1}}{4} \tag{5.21}$$

for suitable k, l.
If we compute the variance at the level below, we obtain

$$\mathrm{Var}(b_{i,j}) = \mathrm{Var}\left(\frac{a_{k,l} + a_{k+1,l} + a_{k,l+1} + a_{k+1,l+1}}{4}\right) = \frac{4\sigma^2}{16} = \frac{\sigma^2}{4}. \tag{5.22}$$

Thus the variance decreases with each level by factor four.

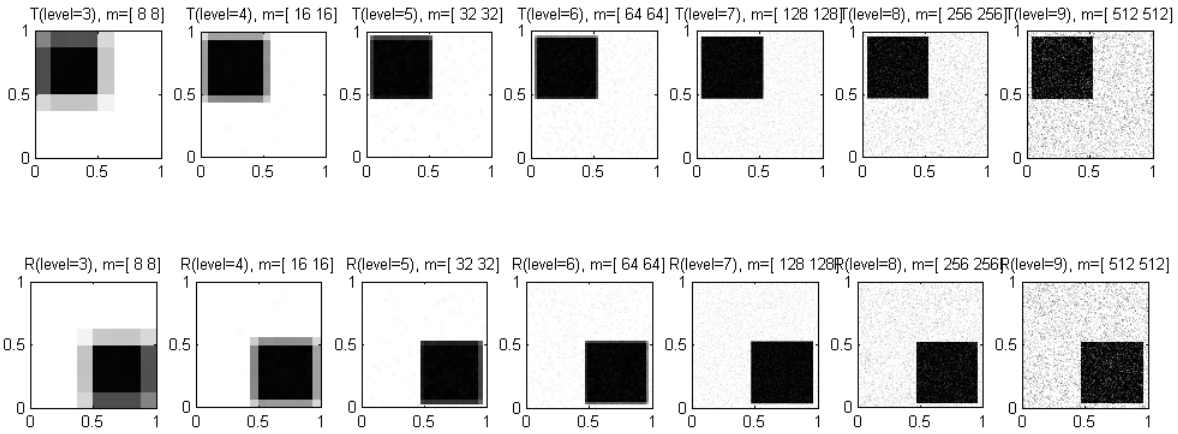


Figure 5.4: Multilevel representation of the registration problem from figure 4.4. Images corrupted by Gaussian noise with mean 0 and standard deviation 50. Note that the noise is nearly completely removed at the lower levels; this suggests to pick smaller regularization parameters on coarser levels.

If we expect our regularization parameter to be linked to the variance of the noise perturbing the data, this suggests to choose smaller regularization weights on coarser levels.
To conclude this example we inspect the difference between the ground truth images (Figure 5.3) and the noise corrupted ones (Figure 5.4):

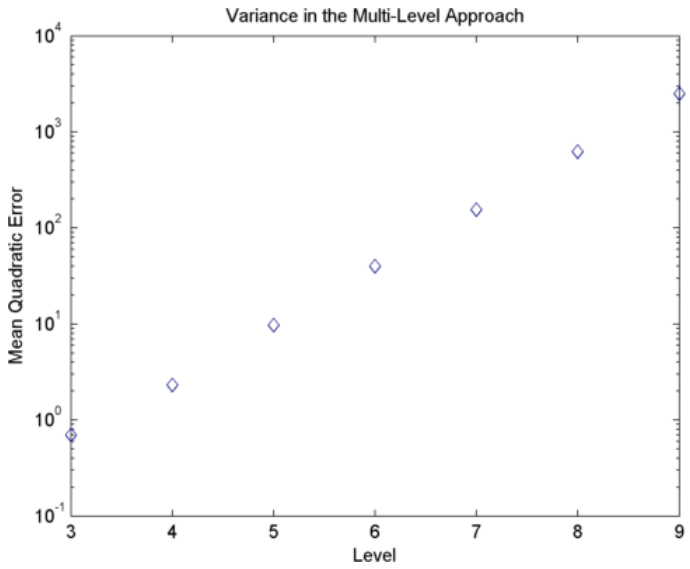


Figure 5.5: Logarithmic plot for the errors produced by the noise in Figure 5.4. The linear increase illustrates (5.22).

This example motivates the use of adaptive regularization techniques, where the parameter is chosen differently on each level. The behaviour of different noise models for the multilevel approach has not yet been examined in detail.

5.2 Optimize-then-Discretize Methods

In this chapter we outline the principles of First-Optimize-then-Discretize methods. The central idea is, to derive optimality conditions for a given functional and then solve them numerically. For deducing these optimality conditions we will present the extension of differential calculus to infinite dimensional spaces. Since these optimality conditions often lead to (weak) solutions of partial differential equations, we present these optimality conditions for functionals of a specific type.

5.2.1 Differentiation for Functionals

In finite dimensions a sufficient (local) optimality condition in $x_0 \in \mathbb{R}^d$ for function $f \in C^2(\mathbb{R}^d; \mathbb{R})$ is given by

$$\nabla f(x_0) = 0 \qquad \nabla^2 f(x_0) \quad \text{positive definite.}$$

For expanding this into infinite dimensions, we will draw our attention on differential calculus on infinite dimensional vector spaces. First we turn our focus to directional derivatives, as introduced by René Gateaux [53]:

Definition 5.2.1 (Gateaux derivative)**.** For any functional $J : \mathcal{U} \to \mathcal{V}$ we define the

directional derivative at u in direction v by

$$d_v J(u) := \lim_{t \downarrow 0} \frac{J(u+tv) - J(u)}{t}. \tag{5.23}$$

Iff this limit exists and is finite, then J is differentiable at u in direction v. The set

$$\delta J(u) := \{d_v J(u) | d_v J(u) < \infty\} \tag{5.24}$$

is called **Gateaux differential**. If $dJ(u) \neq \emptyset$, then J is Gateaux-differentiable at u.

Remarks and Extensions. The directional derivative can also be computed via

$$\frac{d}{dt} J(u+tv)|_{t=0}.$$

The definition from finite dimensional analysis - namely the (total) differential being the linear map approximating the function locally - was carried out towards functionals by Frechet [50, 51].

Definition 5.2.2 (Fréchet differential)**.** Let $\mathcal{U}, \mathcal{V}$ be Banach spaces and $J : \mathcal{U} \to \mathcal{V}$. Furthermore we assume, that the directional derivative $dJ(u; v)$ exists for all directions $v \in \mathcal{U}$. Iff there exists a linear functional $J'(u)$, which fulfills:

$$\langle J'(u), v \rangle = d_v J(u) \tag{5.25}$$

as well as the approximation property

$$\lim_{\|v\|_{\mathcal{U}} \to 0} \frac{\|J(u+v) - J(u) - J'(u)v\|_{\mathcal{V}}}{\|v\|_{\mathcal{U}}} = 0. \tag{5.26}$$

then J is called Fréchet-differentiable at u and $J'(u)$ the **Fréchet differential**.

Analogously $J'(\hat{u}) = 0$ is a necessary condition for $\hat{u}$ being a minimizer of J. We will illustrate this definitions by computing Gateaux and Frechet derivatives for the widely used L_2-distance.

Example 5.2.1 (Derivatives for the (projected) L_2-distance)**.** Let $K : L^2(\Omega) \to L^2(\Omega)$ be a compact operator and $J : L^2(\Omega) \to \mathbb{R}$ be defined via

$$J(U) := \|Ku - f\|_2^2 = \int_\Omega (Ku - f)^2 \mathrm{d}x \tag{5.27}$$

for a suitable domain Ω and given data $f \in L^2(\Omega)$. Since $L^2(\Omega)$ is a Hilbert space this norm can also be written via the L^2 inner product as

$$\|Ku - f\|_2^2 = \langle Ku - f, Ku - f \rangle_{L^2(\Omega)} = \int_\Omega (Ku - f)(Ku - f) \mathrm{d}x. \tag{5.28}$$

We start by computing the directional derivative with the help of formula (5.2.1). Thus we need the derivative

$$\frac{d}{dt}\langle K(u+tv)-f, K(u+tv)-f\rangle. \tag{5.29}$$

For that we expand the dot product:

$$\langle Ku, Ku\rangle + 2t\langle Ku, Kv\rangle + t^2\langle Kv, Kv\rangle - 2\langle Ku, f\rangle - 2t\langle Kv, f\rangle + \langle f, f\rangle. \tag{5.30}$$

Deriving with respect to t leads to

$$\frac{d}{dt}\langle K(u+tv)-f, K(u+tv)-f\rangle = 2t\langle Kv, Kv\rangle + 2\langle Ku-f, Kv\rangle. \tag{5.31}$$

Thus the directional derivative is given by

$$dJ(u;v) = 2\langle Ku-f, Kv\rangle. \tag{5.32}$$

By using the adjoint operator, we expect that the Frechet derivative is given by

$$J'(u) = 2K^*(Ku-f). \tag{5.33}$$

Now only the approximation property (5.26) remains to be shown. We start by evaluating:

$$\begin{aligned}
&\frac{\|J(u+v)-J(u)-J'(u)v\|_2}{\|v\|_2} \\
&= \frac{\|\langle Kv, Kv\rangle\|_2}{\|v\|_2} \\
&= \frac{|\Omega|\|Kv\|_2^2}{\|v\|_2} \\
&\leq \frac{|\Omega|c^2\|v\|_2^2}{\|v\|_2} = |\Omega|c^2\|v\|_2.
\end{aligned}$$

The last inequality holds because K is a compact operator and therefore continuous. Since $|\Omega|c^2\|v\|_2 \to 0$, as $\|v\|_2 \to 0$, the approximation property is verified and J is Frechet differentiable with

$$J'(u) = 2K^*(Ku-f). \tag{5.34}$$

This computation gives us again the normal equation as optimality condition.

5.2.2 Euler-Lagrange Equations

Having computed optimality conditions for a common, yet specific, example occuring in imaging, we turn our focus to a wider class of functionals. According to [45, Chapter 8]

we consider functionals of the type

$$J(v) = \int_\Omega g(Dv(x), v(x), x)\mathrm{d}x = \int_\Omega g(p, z, x)\mathrm{d}x \tag{5.35}$$

with $\Omega \subset \mathbb{R}^d$ and $v : \Omega \to \mathbb{R}^m$ and present the central results Evans [45] states on this subject. In order to characterize minimizers of (5.35) one often varies J around a fixed point $\hat{v}$ in direction v by

$$J(\hat{v} + \tau v) \tag{5.36}$$

for small τ. Then a minimizer $\hat{v}$ of J has to fulfill

$$\frac{d}{d\tau} J(\hat{v} + \tau v)|_{\tau=0} = 0 \tag{5.37}$$

for all $v \in C_c^\infty(\Omega, \mathbb{R}^m)$. Now the condition (5.37) is equivalent to $\hat{v}$ solving

$$-\sum_{i=1}^{d} \frac{\partial}{x_i}\left(\frac{\partial}{\partial p_{k,i}} g(D\hat{v}, \hat{v}, x)\right) + \frac{\partial}{\partial z_k} g(D\hat{v}, \hat{v}, x) = 0 \qquad \text{in } \Omega \qquad (k = 1, ..., m). \tag{5.38}$$

The system of partial differential equations (5.38) is called **Euler-Lagrange equations** for the functional J. If this system is solved by all smooth functions the function L is called a Null-Lagrangian. In this case the energy

$$J(v) = \int_\Omega g(Dv, v, x)\mathrm{d}x \tag{5.39}$$

only depends on boundary values. We will illustrate this definition now with the following example:

Example 5.2.2 (Poisson equation, compare [45])**.** For a smooth function $f : \mathbb{R} \to \mathbb{R}$ we define

$$F(z) := \int_0^z f(x)\mathrm{d}x. \tag{5.40}$$

We consider the energy functional

$$J(w) := \int_\Omega \frac{1}{2}|Dw|^2 - F(w)\mathrm{d}x. \tag{5.41}$$

Consequently we get $\frac{\partial}{\partial p_{k,i}} L = \frac{\partial}{\partial x_i} w$. Thus the Euler-Lagrange equation is given by

$$0 = -\sum_{i=1}^{d} \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_i} w\right) - f(w). \tag{5.42}$$

Carrying out this divergence we end up with the (nonlinear) Poisson equation:

$$-\Delta w = f(w). \tag{5.43}$$

As we saw the minimization of the functional J leads to solving a system of partial differential equations. We will now briefly outline two common strategies how to solve this resulting equations:
In a **finite-difference** framework we discretize the function v on a grid X and compute the differential operators on the grid with finite differences as we presented earlier in this chapter (5.6), (5.7). For example a discrete version of the $2D$-Laplace operator on a grid X with cell size h^2 at the grid point $x_{i,j}$ for a (discrete) function u is given by

$$\Delta^2 u = \frac{u(x_{i+1,j}) + u(x_{i-1,j}) + u(x_{i,j+1}) + u(x_{i,j-1}) - 4u(x_{i,j})}{h^2}. \tag{5.44}$$

With these discrete versions of differential operators, one can derive a (large) system of linear equations, which is solved by the discrete solution. See for example [103] for more details on finite difference methods.
In contrast to that in **finite-element** methods, we triangulate the given domain Ω. On this resulting mesh we can define some basis functions ϕ_l (e.g. spline functions) and get the solution of the differential equation as a linear combination with coefficients c_l of the basis functions, such that these functions fulfill (5.38). For example one could decompose a given domain into small triangles and choose a piecewise linear interpolation on each mesh element as the basis function. See for example [151] for a detailed course to finite element methods.

To conclude this chapter on numerical methods, we shortly state advantages and drawbacks for both optimization approaches. While the Optimize-then-Discretize approach grants a better insight in the structure of the functional to be optimized, this might lead to higher computational costs for solving the optimality conditions [116]. Additionally boundary conditions have to be imposed in order to solve the system of optimality conditions.
In contrast to that this difficulty can be alleviated by using a Discretize-then-Optimize approach: By using differential operators, which switch grid types during the numerical differentiation, the regularizers can be implemented without imposing boundary conditions on the transformation [101]. Furthermore this approach generates a cascade of discrete problems, which converges towards the continuous one. See for example [77, 76] for a detailed discussion of the two approaches.

6 Motion Estimation in Noisy Image Sequences

In this section we present a framework to motion estimation in noisy image sequences. By considering the motion $v \in \mathcal{V}$ as a linear operator

$$T_v u \tag{6.1}$$

acting on an image $u \in \mathcal{U}$, we will deduce a framework for different transformation models and noise characteristics. While we will discuss the spaces $\mathcal{U}$ and $\mathcal{V}$ later, we start with a formulation of the problem and present different choices of transformation models and noise modeling and develop motion estimation functionals for these choices. After discussing challenges in a proper analysis of these functionals we present the numerical realization and some results. We conclude this chapter with a comparison to existing literature on registration with noise modeling and an outlook towards open problems arising in studying this subject.

6.1 Formulation of the Problem

In this section we will derive a framework for motion estimation on noise corrupted image sequences. For the ease of presentation we will limit ourselves to the case of motion estimation between two images. Thus we assume there is an (unknown) underlying true image u and

$$f_0 = T_{v^0} u + \text{"noise"} \qquad f_1 = T_{v^1} u + \text{"noise"} \tag{6.2}$$

are noisy versions of the transformed true image. Note that the noise is not necessarily additive, since we chose the formulation + "noise" to indicate an arbitrary noise corruption. As the noise is of statistical nature, we perform a Bayesian approach to the estimate u, v^0 and v^1 from the noisy data f_0 and f_1. Consequently we aim to maximize the probability

$$p(u, v | f_0, f_1), \tag{6.3}$$

which can be expressed with Bayes' Theorem via

$$p(u, v | f_0, f_1) = \frac{p(f_0, f_1 | u, v) p(u, v)}{p(f_0, f_1)}. \tag{6.4}$$

Analogously to [123, Assumption 2] we assume that noise corrupting our data is context-free, i.e.

$$p(f_0, f_1|u, v) = p_0(f_0|T_{v^0}u)p_1(f_1|T_{v^1}u). \tag{6.5}$$

Now we assume no prior knowledge on the image, thus we set

$$p(u, v) = p(v). \tag{6.6}$$

Ignoring the constant $p(f_0, f_1)$ equation (6.4) evolves now into

$$p_0(f_0|T_{v^0}u)p_1(f_1|T_{v^1}u)p(v). \tag{6.7}$$

We compute a MAP-estimator by minimizing the negative logarithm of (6.7). As usual we set $D_i(u, f_i) := -\log(p_i(f_i|u))$ and $R(v) := -\log(p(v))$. This leads to a minimization problem of the following form:

$$F(u, v) = \int_{\Omega_0} D_0(T_{v^0}u, f_0)\mathrm{d}x + \int_{\Omega} D_1(T_{v^1}u, f_1\mathrm{d}x) + R(v) \tag{6.8}$$

Assuming differentiability of the data fidelities we end up with the following optimality condition for u by using the chain rule:

$$T^*_{v^1}\partial_w D_0(T_{v^0}u, f_0) + T^*_{v^1}\partial_w D_1(T_{v^1}u, f_1) = 0 \tag{6.9}$$

Note that this equation can be solved uniquely for strict convex data fidelities and one injective operator, since we obtain as the second derivative for u:

$$T^*_{v^0}\partial_{ww} D_0(T_{v^0}u, f_0)T_{v^0} + T^*_{v^1}\partial_{ww} D_1(T_{v^1}u, f_1)T_{v^1} \tag{6.10}$$

Let us assume for the sake of clarity in presentation, that T_{v^0} is injective, then for any $x \neq 0$ we directly obtain $T_{v^0}x \neq 0$ and therefore we can deduce by the strict convexity of the D_i

$$\underbrace{x^T T^*_{v^0}\partial_{ww} D_0(T_{v^0}u, f_0)T_{v^0}x}_{>0} + \underbrace{x^T T^*_{v^1}\partial_{ww} D_1(T_{v^1}u, f_1)T_{v^1}x}_{\geq 0}.$$

Having shown that (6.8) is strictly convex with respect to u, it follows that we can solve (6.9) uniquely for u and thus we can define a map

$$U : \mathcal{V} \times \mathcal{U}^2 \to \mathcal{U} \qquad (v; f_0, f_1) \to \overline{u} \tag{6.11}$$

with $\overline{u}$ solving (6.9). Note that $\mathcal{U}$ and $\mathcal{V}$ are appropriate spaces for the given problem. Practically this means $\mathcal{U} = L^p(\Omega)$ with p depending on the data fidelity and $\mathcal{V} = \mathscr{A}$. Consequently the functional to be minimized is obtained by plugging the map into (6.8):

$$J(v) = F(U(v), v) = \int_{\Omega_0} D_0(T_{v^0}, f_0)\mathrm{d}x + \int_{\Omega} D_1(T_{v^1}U(v), f_1)\mathrm{d}x + R(v). \tag{6.12}$$

In the following we will deduce effective functionals for different noise characteristics in the registration framework we presented in Section 4.3.

6.2 Registration Framework for Different Noise Characteristics

In this section we compute different effective functionals for specific noise and transformation models. We use the registration framework we presented in Section 4.3. For a transformation v, which is injective in the sense of Definition 3.3.2, we use the intensity-constant and the mass-preserving transformation operator:

$$T_v^{ic} u := u(v), \qquad T_v^{mp} u := u(v) \det(\nabla v), \tag{6.13}$$

where we can compute the adjoint operators with the area formula as

$$\left(T_v^{ic}\right)^* g = g(v^{-1}) \det(\nabla v^{-1}), \qquad (T_v^{mp})^* g = g(v^{-1}). \tag{6.14}$$

Note that the injectivity of v is crucial, because otherwise we would have the Banach indicatrix as an additional factor in the operator which would cause severe difficulties in the following computations (see Subsection 6.6.4). For the sake of simplicity in the presentation we choose $T_{v^0} = Id$ and T_{v^1} as intensity-constant or mass-preserving transformation operator. Furthermore we use the following identity arising from the implicit function theorem

$$\det(\nabla v^{-1})(v(x)) = \frac{1}{\det(\nabla v)(x)}. \tag{6.15}$$

As different image modalities may cause different noise characteristics we show a few examples in the following table:

Modality	**Noise Type**	**Noise Corruption**	**Data Fidelity**	**Derivative**
Photo Camera	Gaussian	$f = u + \mathcal{N}(0, \sigma)$	$(u - f)^2$	$2(u - f)$
PET/SPECT	Poisson	$f = \mathrm{Poi}(u)$	$u - f \log(u)$	$1 - \frac{f}{u}$
Ultrasound	Speckle	$f = u + \sqrt{u}\mathcal{N}(0, \sigma)$	$\frac{(f-u)^2}{u}$	$\frac{u^2 - f^2}{u^2}$

Table 6.1: Different image modalities with occuring noise characteristics and resulting data fidelites.

The last column gives the derivative of the data fidelity with respect to u. Now we will compute the effective functional for given choices of noise and transformation model. While we carry out these computations for Gaussian noise, computations for Poisson and speckle noise can be found in the Appendix in Section 9.1. For the ease of presentation we assume that u and f_0 are defined on a domain Ω_0 with $\Omega_0 \subset \operatorname{supp}(N_v(\cdot, \Omega))$.

Gaussian Noise

In the case of Gaussian noise, which occurs for example while taking pictures with a photo camera [105], the data fidelity with derivative is given by

$$D(u, f) = (u - f)^2, \qquad \partial_u D(f, u) = 2(u - f).$$

Plugging this into the optimality condition we end up with the equation

$$2(\overline{u} - f_0) + 2T^*_{v^1}(T_{v^1}\overline{u} - f_1) = 0. \tag{6.16}$$

We will solve this equation and derive the effective functional for both presented transformation models.

Intensity-Constant Transformation We solve the optimality condition (6.16) for the intensity-constant trasformation operator:

$$\begin{aligned} 0 &= 2(\overline{u} - f_0) + 2\left(T^{ic}_{v^1}\right)^*(\overline{u}(v) - f_1) \\ 0 &= 2(\overline{u} - f_0) + 2\det(\nabla v^{-1})(\overline{u} - f_1(v^{-1})) \\ 2(1 + \det(\nabla v^{-1}))\overline{u} &= 2f_0 + 2f_1(v^{-1})\det(\nabla v^{-1}) \\ \overline{u} &= \frac{f_0 + f_1(v^{-1})\det(\nabla v^{-1})}{1 + \det(\nabla v^{-1})} \end{aligned}$$

Now applying the transformation operator leads to

$$\overline{u}(v) = \frac{f_0(v) + f_1 \frac{1}{\det(\nabla v)}}{1 + \frac{1}{\det(\nabla v)}} = \frac{f_0(v)\det(\nabla v) + f_1}{1 + \det(\nabla v)}.$$

Having computed the map we plug this into 6.12 and examine both data fidelities for f_0 and f_1 terms separately. The first term leads to

$$\begin{aligned} D(\overline{u}, f_0) &= \int_{\Omega_0} (\overline{u}(x) - f_0(x))^2 \mathrm{d}x \\ &= \int_{\Omega} (\overline{u}(v(x)) - f_0(v(x)))^2 \det(\nabla v)(x) \mathrm{d}x \end{aligned}$$

$$= \int_\Omega \left(\frac{f_0(v(x))\det(\nabla v) + f_1(x)}{1+\det(\nabla v)} - f_0(x) \right)^2 \det(\nabla v)\mathrm{d}x$$
$$= \int_\Omega \left(\frac{f_1(x) - f_0(v(x))}{1+\det(\nabla v)} \right)^2 \det(\nabla v)\mathrm{d}x,$$

while the second computation is given by

$$\begin{aligned} D(T_{v^1}\overline{u}, f_1) &= \int_\Omega (\overline{u}(v(x)) - f_1(x))^2 \mathrm{d}x \\ &= \int_\Omega \left(\frac{f_0(v(x))\det(\nabla v(x)) + f_1(x)}{1+\det(\nabla v(x))} - f_1 \right)^2 \mathrm{d}x \\ &= \int_\Omega \left(\frac{f_0(v(x))\det(\nabla v(x)) - f_1(x)\det(\nabla v(x))}{1+\det(\nabla v)} \right)^2 \mathrm{d}x. \end{aligned}$$

Adding this we end up with

$$\begin{aligned} D^{gaussian,ic} &= D(\overline{u}, f_0) + D(T_{v^1}\overline{u}, f_1) \\ &= \int_\Omega \left(\frac{f_1(x) - f_0(v(x))}{1+\det(\nabla v)} \right)^2 \det(\nabla v) \\ &\quad + \left(\frac{f_0(v(x))\det(\nabla v(x)) - f_1(x)\det(\nabla v(x))}{1+\det(\nabla v)} \right)^2 \mathrm{d}x \\ &= \int_\Omega \left(\frac{f_0(v(x)) - f_1(x)}{1+\det(\nabla v(x))} \right)^2 \det(\nabla v(x)) \\ &\quad + \left(\frac{f_0(v(x)) - f_1(x)}{1+\det(\nabla v(x))} \right)^2 \det(\nabla v(x))^2 \mathrm{d}x \\ &= \int_\Omega \left(\frac{f_0(v(x)) - f_1(x)}{1+\det(\nabla v(x))} \right)^2 \det(\nabla v(x))(1+\det(\nabla v(x)))\mathrm{d}x \\ &= \int_\Omega (f_0(v(x)) - f_1(x))^2 \frac{\det(\nabla v(x))}{1+\det(\nabla v(x))}\mathrm{d}x. \end{aligned}$$

The result is now closely related to the classical SSD-distance but with the additional weighting term $\frac{\det(\nabla v(x))}{1+\det(\nabla v(x))}$.

Mass-Preserving Transformation We solve the optimality condition (6.16) for the mass-preserving transformation operator:

$$\begin{aligned}
0 &= 2(\overline{u} - f_0) + 2\,(T_v^{mp})^*\,(\overline{u}\det(\nabla v) - f_1) \\
0 &= (\overline{u} - f_0) + (\overline{u}\frac{1}{\det(\nabla v^{-1})} - f_1(v^{-1})) \\
\overline{u}\left(1 + \frac{1}{\det(\nabla v^{-1})}\right) &= f_0 + f_1(v^{-1}) \\
\overline{u} &= \frac{f_0 + f_1(v^{-1})}{1 + \frac{1}{\det(\nabla v^{-1})}} \\
\overline{u} &= \frac{f_0\det(\nabla v^{-1}) + f_1(v^{-1})\det(\nabla v^{-1})}{1 + \det(\nabla v^{-1})}.
\end{aligned}$$

Having computed the map for any fixed transformation v, we now compute the composition of said map with the transformation v

$$\overline{u}(v) = \frac{f_0(v)\frac{1}{\det(\nabla v)} + f_1\frac{1}{\det(\nabla v)}}{1 + \frac{1}{\det(\nabla v)}} = \frac{f_0(v) + f_1}{1 + \det(\nabla v)}.$$

With this at hand we can compute both data fidelities:

$$\begin{aligned}
D_0(\overline{u}, f_0) &= \int_{\Omega_0} (\overline{u}(x) - f_0(x))^2 \mathrm{d}x \\
&= \int_{\Omega} (\overline{u}(v(x)) - f_0(v(x)))^2 \det(\nabla v)\mathrm{d}x \\
&= \int_{\Omega} \left(\frac{f_0(v(x)) + f_1(x)}{1 + \det(\nabla v)} - f_0\right)^2 \det(\nabla v)\mathrm{d}x \\
&= \int_{\Omega} \left(\frac{f_1(x) - f_0(v(x))\det(\nabla v)}{1 + \det(\nabla v)}\right)^2 \det(\nabla v)\mathrm{d}x \\
&= \int_{\Omega} \left(\frac{f_0(v(x))\det(\nabla v) - f_1(x)}{1 + \det(\nabla v)}\right)^2 \det(\nabla v)\mathrm{d}x.
\end{aligned}$$

For the second data fidelity we end up with:

$$\begin{aligned}
D_1(T_v^{mp}\overline{u}, f_1) &= \int_{\Omega} \left(\overline{u}(v(x))\det(\nabla v) - f_1(x)\right)^2 \mathrm{d}x \\
&= \int_{\Omega} \left(\frac{f_0(v(x))\det(\nabla v) + f_1\det(\nabla v)}{1 + \det(\nabla v)} - f_1\right)^2 \mathrm{d}x
\end{aligned}$$

$$= \int_\Omega \left(\frac{f_0(v(x))\det(\nabla v) - f_1(x)}{1+\det(\nabla v)} \right)^2 \mathrm{d}x.$$

The final distance measure is now given as the sum of both data fidelities:

$$\begin{aligned} D^{gaussian,mp} &= D(\overline{u}, f_0) + D(T_v^{mp}\overline{u}, f_1) \\ &= \int_\Omega \left(\frac{f_0(v(x)\det(\nabla v) - f_1(x)}{1+\det(\nabla v)} \right)^2 (1+\det(\nabla v))\mathrm{d}x \\ &= \int_\Omega \left(f_0(v(x)\det(\nabla v) - f_1(x)\right)^2 \frac{1}{1+\det(\nabla v)}\mathrm{d}x. \end{aligned}$$

This is the SSD-distance for the mass-preserving transformed template with the weighting $\frac{1}{1+\det(\nabla v)}$.
For the sake of completeness we shortly list the distance measures for other noise types; see Section 9.1 in the Appendix for detailed computation.

$$\begin{aligned} D^{poisson,ic} &= \int_\Omega f_0(v(x))\det(\nabla v)\ln(f_0(v)) \\ &\qquad - (f_0(v(x))\det(\nabla v) + f_1(x))\ln\left(\frac{f_0(v(x))\det(\nabla v) + f_1(x)}{1+\det(\nabla v)}\right)\mathrm{d}x \\ D^{poisson,mp} &= \int_\Omega f_0(v)\det(\nabla v)\ln(f_0(v)\det(\nabla v)) + f_1\ln(f_1) \\ &\qquad - (f_0(v)\det(\nabla v) + f_1)\ln\left(\frac{f_0(v)\det(\nabla v) + f_1}{2}\right)\mathrm{d}x \\ D^{speckle,ic} &= \int_\Omega 2\sqrt{(f_0(v)^2\det(\nabla v) + f_1^2)(1+\det(\nabla v))} - 2(f_0(v)\det(\nabla v) + f_1)\mathrm{d}x \\ D^{speckle,mp} &= \int_\Omega \frac{2(\overline{u}(v)\det(\nabla v))^2 + (f_0(v)\det(\nabla v))^2 + f_1^2}{\overline{u}(v)\det(\nabla v)} - 2(f_0(v)\det(\nabla v) + f_1)\mathrm{d}x \end{aligned}$$

6.3 Analysis

In this section we present detailed analysis for the resulting functionals in the registration framework. We start by describing the regularization properties and motivate the choice of hyperelastic regularization.

6.3.1 Regularization

Since all the functionals depend on the Jacobian determinant, which can be traced back to the use of the adjoint operator for the intensity-constant transformation model,

we decide to use hyperelastic regularization. Additionally this control of the Jacobian determinant grants us at least local injectivity of the transformation. This can be used to force global injectivity via controlling the behaviour of the function on the boundary (see Proposition 3.2.9).
As we have described different regularization types in the previous chapters, we just present the used hyperelastic energy R_{hyper}:

$$R_{\text{hyper}}(v) \;=\; \int_\Omega \alpha_1 \operatorname{len}(\nabla v) \;+\; \alpha_2 \operatorname{surf}(\operatorname{cof}(\nabla v)) \;+\; \alpha_3 \operatorname{vol}(\det(\nabla v))\,\mathrm{d}x \;,$$

with $\alpha_i \geq 0, \sum \alpha_i > 0$ for $i = 1, 2, 3$ and the penalty functions

$$\operatorname{len}(s) = \|s - \mathcal{I}\|^2_{\text{Fro}} \;, \qquad \operatorname{surf}(s) = \left(\|s\|^2_{\text{Fro}} - 3\right)^2 , \qquad \operatorname{vol}(s) = \frac{(s-1)^4}{s^2} \;.$$

6.3.2 Existence of Minimizers

In this section we discuss difficulties in showing the existence of minimizers for the effective functionals with noise modeling. As we recall the direct method in the calculus of variations we need to verify for each functional:

(ls) lower semicontinuity

(csls) compactness of sublevel sets

While we can deduce the compactness of the sublevel sets by coercivity properties of the regularizer and the Banach-Alaoglu theorem (Theorem 3.4.7), the lower semicontinuity might cause severe problems.
For a functional of the type

$$J(v) = \int_\Omega g(Dv(x), v(x), x)\mathrm{d}x \tag{6.17}$$

the following theorem from Evans ([45]) gives a sufficient condition for weak lower semicontinuity:

Theorem 6.3.1 (Weak lower semicontinuity)**.** *Assume that g is bounded below, and in addition the mapping*

$$p \to g(p, z, x) \tag{6.18}$$

is convex for each $z, x \in \Omega$. Then J is weakly lower semicontinuous.

Proof. See [45, Chapter 8, Theorem 1]. □

Generalization is given by Theorems 8 and 9. Consequently we investigate, if the resulting effective functionals are convex with respect to the Jacobian determinant. As it turns out, this is not always the case.

Distance Measure	Second derivative	Convexity
$D^{gaussian,ic}$	$\frac{-2(f_0(v)-f_1)^2}{(\det(\nabla v)+1)^3}$	concave
$D^{gaussian,mp}$	$2\frac{(f_0(v)+f_1)^2}{(1+\det(\nabla v))^3}$	convex
$D^{poisson,ic}$	$-\frac{(f_0(v)-f_1)^2}{(\det(\nabla v)+1)^2(f_0(v)\det(\nabla v)+f_1)}$	concave
$D^{poisson,mp}$	$\frac{f_0(v)f_1}{\det(\nabla v)(f_0(v)\det(\nabla v)+f_1)}$	convex
$D^{speckle,ic}$	$-\frac{(f_0(v)^2-f_1^2)^2}{4\sqrt[3]{(f_0(v)^2\det(\nabla v)^2+(f_0(v)^2+f_1^2)\det(\nabla v)+f_1^2)^2}}$	concave
$D^{speckle,mp}$	$\frac{f_0(v)^2f_1^2}{\sqrt[3]{(f_0(v)^2\det(\nabla v)^2+f_1^2)^2}}$	convex

Table 6.2: The second derivative with respect to the Jacobian is a sufficient criterion to evaluate (poly)convexity.

We see in the table, that each functional for the mass-preserving transformation operator is polyconvex. In fact we can even generalize this observation with the following theorem:

Theorem 6.3.2. *For strict convex data fidelities $D_1, D_2 \in C^2(D)$ with a suitable domain $D \subset \mathbb{R}$, let J be the functional defined by* (6.12) *for the mass-preserving transformation operator. Then J is polyconvex.*

Proof. After using the area formula for the first distance term the effective functional is of the following type:

$$J(v) = \int_\Omega \underbrace{D_0(u(v))\det(\nabla v) + D_1(u(v)\det(\nabla v))}_{:=j(\det(\nabla v))}\,\mathrm{d}x. \tag{6.19}$$

With the abbreviations $\tilde{u} := u(v)$ and $d := \det(\nabla v)$ this reads as

$$\int_\Omega \underbrace{D_0(\tilde{u}(d))d + D_1(\tilde{u}(d)d)}_{:=j(d)}\,\mathrm{d}x.$$

For this functional we can give a transformed first order optimality condition like (6.9) for $\tilde{u}$:

$$D_0'(\tilde{u})d + D_1'(\tilde{u}d) = 0. \tag{6.20}$$

Note that for any fixed v u fulfills (6.9), whenever $\tilde{u}$ fulfills (6.20).
With this at hand we prove convexity by deriving $j(d)$ again with the abbreviation $\tilde{u} := \tilde{u}(d)$:

$$j'(d) = \tilde{u}D_0'(\tilde{u})d + D_0(\tilde{u}) + (\tilde{u}'d + \tilde{u})D_1'(\tilde{u}d) \tag{6.21}$$

This leads to the second derivative:

$$\begin{aligned} j''(d) =& \tilde{u}''D_0'(\tilde{u})d + (\tilde{u}')^2 D_0''(\tilde{u})d + 2\tilde{u}'D_0'(\tilde{u}) \\ &+ (\tilde{u}''d + 2\tilde{u}')D_1'(\tilde{u}d) + (\tilde{u}'d + \tilde{u})^2 D_1''(\tilde{u}d) \end{aligned}$$

We rearrange this term for more clarity:

$$j''(d) = (\tilde{u}'' + 2\tilde{u}') \underbrace{(D_0'(\tilde{u})d + D_1'(\tilde{u}d)d)}_{:=I} + \underbrace{(\tilde{u}')^2 D_0''(\tilde{u})d + (\tilde{u}'d\tilde{u})^2 + D_1''(\tilde{u}d)}_{:=II} \tag{6.22}$$

While the first term I vanishes because of the transformed optimality condition (6.20), the second is non-negative because of the strict convexity of the data fidelities. This is sufficient for convexity in the Jacobian determinant and thus polyconvexity.

□

Having shown the semicontinuity for a wide class of transformations, we can use Theorem 6.3.1 for a general existence result for the mass-preserving transformation operator:

Theorem 6.3.3 (Existence of a minimizer for mass-preserving transformations)**.** *Let $D_1, D_2 \in C^2(D)$ be strict convex data fidelities for a suitable domain $D \subset \mathbb{R}$. Then the registration functional with noise modeling defined by* (6.12) *for the mass-preserving transformation operator has a minimizer.*

Proof. We use the direct method in the calculus of variations, thus we need to verify lower semicontinuity and compactness of sublevel sets.
Since we proved polyconvexity of the functional in Theorem 6.3.2, Theorem 6.3.1 directly grants us lower semicontinuity.
Furthermore, we can show coercivity directly from the properties of the hyperelastic regularization, since the strict convexity of the data fidelities implies a lower bound for the distance measure. Then Banach-Alaoglu gives compactness of the sublevel sets. □

For the intensity-constant transformation we observe that each of the deduced distance measures is concave with respect to the Jacobian determinant. In fact we can prove this property for a wide class of data fidelities:

Theorem 6.3.4. *For strict convex data fidelities $D_1, D_2 \in C^2(D)$, where D is a suitable domain for the fidelity, let J be the functional defined by* (6.12) *for the intensity-constant transformation operator. Then J is concave with respect to the Jacobian determinant.*

Proof. Again we start by using the area formula for the first data fidelity and end up with

$$J(v) = \int_\Omega \underbrace{D_0(u(v)) \det(\nabla v) + D_1(u(v))}_{:= j(\det(\nabla v))} \mathrm{d}x. \tag{6.23}$$

With the same abbreviations as above we get

$$J(v) = \int_\Omega \underbrace{D_0(\tilde{u}(d))d + D_1(\tilde{u}(d))}_{:= j(d)} \mathrm{d}x.$$

Again we derive a transformed version of the optimality condition, which is again fulfilled, whenever (6.9) holds for any fixed v

$$D_0'(\tilde{u})d + D_1'(\tilde{u}) = 0. \tag{6.24}$$

For the later parts of the proof we differentiate this optimality condition with respect to d

$$\tilde{u}' D_0''(\tilde{u})d + D_0'(\tilde{u}) + \tilde{u}' D_1''(\tilde{u}) = 0. \tag{6.25}$$

With this preparations we proceed in a similar way as in the theorem above and compute the derivatives:

$$j'(d) = \tilde{u}' D_0'(\tilde{u})d + D_0(\tilde{u}) + \tilde{u}' D_1''(\tilde{u}). \tag{6.26}$$

This leads again to the second derivative:

$$j''(d) = \tilde{u}'' D_0(\tilde{u})d + 2\tilde{u}' D_0'(\tilde{u}) + (\tilde{u})^2 D_0''(\tilde{u})d + \tilde{u}'' D_1'(\tilde{u}) + (\tilde{u}')^2 D_1''(\tilde{u}) \tag{6.27}$$

Now we deduce the following identity from (6.25)

$$D_0'(\tilde{u}) = -\tilde{u}' D_0''(\tilde{u})d - \tilde{u}' D_1''(\tilde{u}). \tag{6.28}$$

Plugging this into the second derivative we end up with:

$$j''(d) = \underbrace{\tilde{u}''(D_0'(\tilde{u})d + D_1'(\tilde{u}))}_{=0} \underbrace{-((\tilde{u}')^2(D_0''(\tilde{u})d + D_1''(\tilde{u})))}_{\leq 0}. \tag{6.29}$$

Again the first term vanishes because of the first order optimality condition (6.24), while the second one is non-positive because of the strict convexity of the data fidelities.

□

Unfortunately, this shows that we do not have the chance to use the direct method in the calculus of variations for the intensity constant transformation operator. Even worse, we can directly deduce upper semicontinuity for the distance measures by using the polyconvexity of the negative distance measure, which is convex with respect to the Jacobian determinant. Therefore we cannot expect lower semicontinuity, since functionals depending on the Jacobian determinant are in general not continuous. Additionally, because of the fact that the distance measure is concave with respect to the Jacobian determinant, convexification will not work.
Thus we need another approach to tackle this problem. As a first idea we go back to the first formulation of the functional (6.8) and reformulate this problem:

$$\begin{aligned} \min_{u,u_0,u_1,v} \quad & D_0(u_0,f_0) + D_1(u_1,f_1) + R(v) \\ \text{such that} \quad & T_0(v)u = u_0 \\ & T_{v^1}u = u_1. \end{aligned}$$

As we can observe the minima of the original and the reformulated functional coincide. Thus we can benefit from the simplified structure of the functional, at the prize of dealing with a side condition. Ignoring this condition for the moment, we can show (csls) and (ls) from the properties of the distance measures. But the prize we have to pay is that we need to show the closedness of the constraint.
As mentioned before each admissible transformation v defines a linear operator

$$T_v^{ic} : L_2 \to L_2 \qquad u \to u(v). \tag{6.30}$$

Lemma 6.3.5. *Let v be an admissible transformation with $0 < \delta \leq \det(\nabla v)$. Then the corresponding transformation operator is continuous.*

Proof.

$$\|T_v^{ic}u\|_2^2 = \int_\Omega (u(v(x)))^2 \,\mathrm{d}x = \int_\Omega (u(x)\det(\nabla v^{-1}))^2 \,\mathrm{d}x \leq \frac{1}{\delta^2}\int_\Omega u(x)^2 \,\mathrm{d}x = \frac{1}{\delta^2}\|u\|_2^2$$

□

As a consequence we directly obtain $\|T_v^{ic}\| \leq \frac{1}{\delta}$. With that at hand we can investigate convergence properties of the corresponding operator, given a weakly convergent sequence of admissible transformations:

Lemma 6.3.6. *Let $v_n \rightharpoonup v$ in H^1, such that $v_n, v \in \mathscr{A}$. Then the sequence of corresponding transformation operators T_{v_n} converges to T_v in the strong operator topology.*

Proof. Since $\|T_{v_n}\| \leq \frac{1}{\delta} < \infty$ it is sufficient to show the convergence on a dense subset of L_2: Let $\tilde{u}$ be a Lipschitz function with Lipschitz-constant L, then we obtain:

$$\|\tilde{u}(v) - \tilde{u}(v_n)\| \leq L^2\|v - v_n\|_2^2$$

Since weak convergence in H^1 implies strong L_2-convergence, the convergence is shown.
□

With these preparations at hand we can formulate a convergence result:

Theorem 6.3.7. *Let $x_n \to x$, $u_n \to u$ in L_2 and $v_n \rightharpoonup v$ in $\mathscr{A}$ with*

$$T_{v_n} u_n = x_n, \tag{6.31}$$

then this property also holds in the limit, i.e.

$$T_v u = x. \tag{6.32}$$

Proof. We show for any $\epsilon > 0$:

$$\|T_v u - x\| < \epsilon.$$

For $\xi \in L_2$ with $\|\xi - u\| < \delta \frac{\epsilon}{8}$ we use the following estimation:

$$\|T_v u - x\| \leq \underbrace{\|T_v u - T_v \xi\|}_{:=I} + \underbrace{T_v \xi - T_{v_n} \xi\|}_{:=II} + \underbrace{\|T_{v_n} \xi - T_{v_n} u_n\|}_{:=III} + \underbrace{\|T_{v_n} u_n - x\|}_{:=IV}$$

We examine these four parts separately:
I: We obtain:

$$\|T_v u - T_v \xi\| \leq \frac{1}{\delta} \|x - \xi\| = \frac{1}{\delta} \delta \frac{\epsilon}{8} = \frac{\epsilon}{8}.$$

II: Since T_{v_n} converges to T_v in the strong operator topology, we find a N_1 such that:

$$\|T_v \xi - T_{v_n} \xi\| < \frac{\epsilon}{8} \quad \forall n \geq N_1.$$

III: By using the continuity of the transformation operator we obtain:

$$\|T_{v_n} \xi - T_{v_n} u_n\| \leq \frac{1}{\delta} \|\xi - u_n\| \leq \frac{1}{\delta} (\|\xi - u\| + \|u_n - u\|).$$

Since $u_n \to u$ we find N_2 with

$$\|u_n - u\| < \delta \frac{\epsilon}{8}, \qquad \|T_{v_n} \xi - T_{v_n} u_n\| < \frac{1}{\delta} (\frac{\delta \epsilon}{8} + \frac{\delta \epsilon}{8}) = \frac{\epsilon}{4} \quad \forall n \geq N_2.$$

IV: By property (6.31) we can deduce:

$$\|T_{v_n} u_n - x\| = \|x_n - x\| < \frac{\epsilon}{4} \quad \forall n \geq N_3$$

for some $N_3 \in \mathbb{N}$, since x_n converges strongly to x.
If we choose any $n \geq N := \max\{N_1, N_2, N_3\}$ the assertion follows.

□

Note that for this theorem the strong convergence is crucial. If this is not given we end up with a product of two weakly convergent sequences, where in general no statement can be given. Unfortunately strong convergence can only be guaranteed with further regularity assumptions for the image u, which would complicate the first order optimality condition (6.9), so this is not an option.

6.4 Numerical Framework

In this section we present the numerical framework, which we use to minimize the registration functionals, which we derived for different noise characteristics. As we saw these were in fact registration functionals, but with some new distance functions. We will modify the FAIR-framework in order to deal with these distance functions. Our framework follows a First-Discretize-then-Optimize approach, like we described in Chapter 5. For a given template and transformation grid, the transformed template is computed by interpolation (see ([101] for more details). The difficulty is caused by the fact that all the functionals depend on the Jacobian determinant $\det(\nabla v)$ of the transformation. We start by describing the (discrete) notation we will use in this section:

6.4.1 Implementation of an Objective Function

We assume that we have two digital images f_0, f_1 with $m = m_1 \times ... \times m_d$ pixels, such that

$$f_0, f_1 \in \mathbb{R}^m. \tag{6.33}$$

The (discrete) transformation is then given as a nodal transformation grid

$$vc \in \mathbb{R}^{\tilde{m}} \qquad \tilde{m} = \prod_{i=1}^{d} (m_i + 1)d. \tag{6.34}$$

Furthermore we denote by

$$\text{inter}(f_0, vc) : \mathbb{R}^m \times \mathbb{R}^{\tilde{m}} \to \mathbb{R}^m \tag{6.35}$$

the transformation of f_0 with the motion vectors vc by interpolation. The computation of the Jacobian determinant is given by

$$\text{jac}(vc) : \mathbb{R}^{\tilde{m}} \to \mathbb{R}^m. \tag{6.36}$$

As a consequence of the chain rule we can directly compute the derivatives with respect to vc for continuously differentiable distance measures

$$D : (\mathbb{R}^m)^3 \to \mathbb{R} \qquad (\text{inter}(f_0, vc), \text{jac}(vc), f_1) \to D(\text{inter}(f_0, vc), \text{jac}(vc), f_1) \tag{6.37}$$

as

$$\frac{\mathrm{d}}{\mathrm{d}\, vc} D(\text{inter}(f_0, vc), \text{jac}(vc), f_1) = \frac{\mathrm{d}}{\mathrm{d}\, w_1} D \frac{\mathrm{d}}{\mathrm{d}\, vc} \text{inter} + \frac{\mathrm{d}}{\mathrm{d}\, w_2} D \frac{\mathrm{d}}{\mathrm{d}\, vc} \text{jac}\,. \tag{6.38}$$

In order to perform a classic matrix-vector multiplication, we compute the transpose of (6.38) in our framework.
Computing the derivative with help of (6.38) has the consequence that we only have to differentiate the distance term with respect to the transformed template and the

Jacobian determinant, in order to compute the derivative with respect to the transformation. To conclude the course on distance measures we present a short synopsis of all distance measures with their derivatives. We start by derivating the distance measures with respect to the transformed template:

Distance	**Derivative with respect to $f_0(v)$**
$D^{gaussian,ic}$	$2(f_0(v) - f_1)\frac{\det(\nabla v)}{1+\det(\nabla v)}$
$D^{gaussian,mp}$	$2\frac{\det(\nabla v)(f_0(v)\det(\nabla v)-f_1)}{1+\det(\nabla v)}$
$D^{poisson,ic}$	$\det(\nabla v)\left(\ln(f_0(v)) - \ln\left(\frac{f_0(v)\det(\nabla v)+f_1}{1+\det(\nabla v)}\right)\right)$
$D^{poisson,mp}$	$\det(\nabla v)\ln(f_0(v)\det(\nabla v)) - \det(\nabla v)\ln\left(\frac{f_0(v)\det(\nabla v)+f_1}{2}\right)$
$D^{speckle,ic}$	$\frac{f_0(v)\det(\nabla v)(1+\det(\nabla v))}{\sqrt{(f_0(v)^2\det(\nabla v)+f_1)(1+\det(\nabla v))}} - 2\det(\nabla v)$
$D^{speckle,mp}$	$\frac{4f_0(v)\det(\nabla v)^2}{\sqrt{2(f_0(v)^2\det(\nabla v)^2+f_1^2)}} - 2\det(\nabla v)$

Table 6.3: Derivatives of the continuous integrand functions with respect to the transformed template. The discrete derivative can be computed with help of the discretization of the distance measure, since the discrete integral is a sum of the integration function, evaluated on grid points multiplied by the cell size.

Having computed the continuous derivatives with respect to the transformed template $f_0(v)$, the derivatives of a discretized objective function follow immediately from the nature of the chosen discretization. To complete the overview on the derivatives for our distance measures, we just state the derivatives of the distance measures with respect to the Jacobian determinant $\det(\nabla v)$:

Distance	**Derivative for** $\det(\nabla v)$
$D^{gaussian,ic}$	$(f_0(v) - f_1)^2 \frac{1}{(1+\det(\nabla v))^2}$
$D^{gaussian,mp}$	$-\frac{2f_0(v)(f_0(v)\det(\nabla v)-f_1)}{(1+\det(\nabla v))^2}$
$D^{poisson,ic}$	$-f_0(v)\ln\left(\frac{f_0(v)\det(\nabla v)+f_1}{1+\det(\nabla v)}\right) - f_0(v)(\ln(f_0(v)) - 1) - \frac{f_0(v)\det(\nabla v)+f_1}{\det(\nabla v)+1}$
$D^{poisson,mp}$	$f_0(v)\ln(f_0(v)\det(\nabla v)) - f_0(v)\ln\left(\frac{f_0(v)\det(\nabla v)+f_1}{2}\right)$
$D^{speckle,ic}$	$\frac{f_0(v)^2(1+2\det(\nabla v))}{\sqrt{(f_0(v)^2\det(\nabla v)+f_1)(1+\det(\nabla v))}} - 2f_0(v)$
$D^{speckle,mp}$	$\frac{4f_0(v)^2\det(\nabla v)}{\sqrt{2(f_0(v)^2\det(\nabla v)^2+f_1^2)}} - 2f_0(v)$

Table 6.4: Derivatives of the continuous integrand functions with respect to the Jacobian determinant. Again the derivatives of the discretized objective function can be computed in a straightforward manner.

The objective function can be implemented by using the routines from the FAIR toolbox [101] for interpolation, regularization and computation of the Jacobian determinant. See Section 9.2.1 in the Appendix.

6.4.2 Numerical Minimization

We minimize the objective function according to the framework presented in Chapter 5. Thus we follow the multilevel approach, proposed in Section 5.1.3. This leads to the minimization of a discrete problem on each level, where the solution on a coarse level serves as initial guess for the finer level.
To solve this optimization problem on each level, we use the modified BFGS method, presented in Chapter 5. Since we follow the modular implementation of the FAIR toolbox [101], we can modify the multilevel approach, provided by FAIR to fit our needs. Thus our minimization algorithm reads as follows:

1. Obtain multilevel representations for the two d-dimensional images f_0 and f_1 to be registered and choose a regularization parameter α.

2. Choose a minimum level l_{min} and a maximum level l_{max} for the multilevel framework.

3. Perform a parametric pre-registration. Set the result as starting value.

4. Set $j = l_{min}$

5. Change the weight of the regularization to $\alpha(2^d)^{-(l_{max}-j)}$.

6. Solve the problem on the specified j-th level with the given start value.

7. If $j < l_{max}$ prolong the solution to the finer level and set $j = j + 1$, else return the solution and end.

8. Repeat steps 5-7.

Note that the adaptive choice of the regularization parameter is motivated by Example 5.1.1. An implementation of this algorithm can be found in Section 9.2.1 in the Appendix.

6.5 Results

In this section we compare the proposed MAP-estimation with results from the classical (mass-preserving) registration on (presmoothed) noisy data. Before we focus on some test problems, we shortly describe how we assess the quality of the estimated transformation. Note that often no ground truth vectors, but only ground truth images u_1, u_2 are available. Even if ground truth motion vectors are available computing the norm of the difference between ground truth and estimated motion v might be critical, since different transformations can produce a perfect registration result (compare Example 4.3.1). So instead we compare how the transformations behave on the ground truth images. This idea is captured by the phantom-matching-error (compare [96])

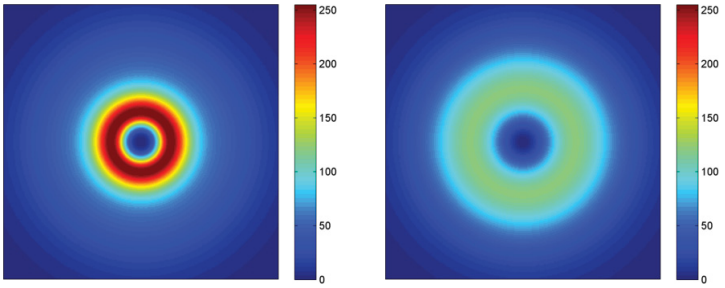
$$\mathrm{PME}(v; u_1, u_2) := \int_\Omega (T(v)u_1 - u_2)^2 \mathrm{d}x. \tag{6.39}$$

The quality of transformations providing similar phantom matching errors can be assessed by comparison of the regularization energy. Additional to these analytical approaches we use visual inspection to evaluate the results provided by the methods above: This includes plotting the transformation grids, the norm of the displacement field and the Jacobian determinant for a visual impression of the regularity as well as showing the transformed ground truth images in order to highlight the accuracy of the transformed template, which is measured by the phantom-matching error.
We start the presentation of results with a synthetic example for mass-preserving registration. Note that this example was already presented in [142]. By adding noise with different variance to the ground truth images we are able to compare our MAP-estimation with a classical registration and the aligning of presmoothed images. We

restrict our presentation to Gaussian noise, but results for Poisson noise and cross tests for the noise models are also available [142]. We conclude this Section by applying the MAP-estimation on a realistic software phantom in order to show the potential of the method in practical applications.

6.5.1 Synthetic Toy Example

We start with a simple, but yet challenging registration task:



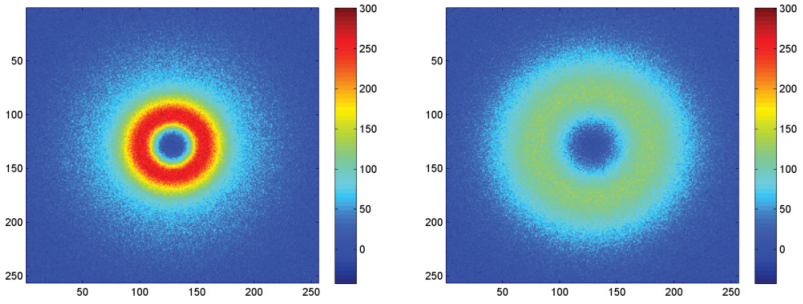
(a) Reference (b) Template

Figure 6.1: Ground truth data for mass preserving registration: intensity distributed over a wide area (template) and the same amount concentrated in a small circle (reference).

Since intensity needs to be changed for a perfect registration, we choose the mass-preserving transformation operator. For different noise levels we compare the proposed method with a classical mass-preserving registration performed by the VAMPIRE app [63] of the FAIR toolbox [101] performed on the noisy and on presmoothed data. In order to highlight the strength of the proposed method we compare these registration methods for different noise levels.

Low Noise

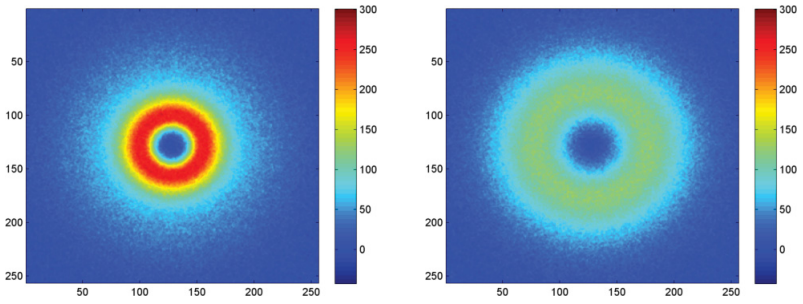
In the case of a low noise corruption the images to be registered are given by:



(a) Noisy reference (b) Noisy template

Figure 6.2: Noisy images to registered. Images corrupted by Gaussian noise with standard deviation $\sigma = 10$.

Applying a Gaussian filter smoothes the images and we get:



(a) Filtered reference (b) Filtered template

Figure 6.3: Noisy images in Figure 6.2 filtered with a Gaussian kernel.

In order to assess the quality of the transformations estimated by the different registration methods we inspect the phantom matching errors for different hyperelastic regularization parameters $\alpha_1 = \alpha_3 = \alpha$ and $\alpha_2 = 0$:

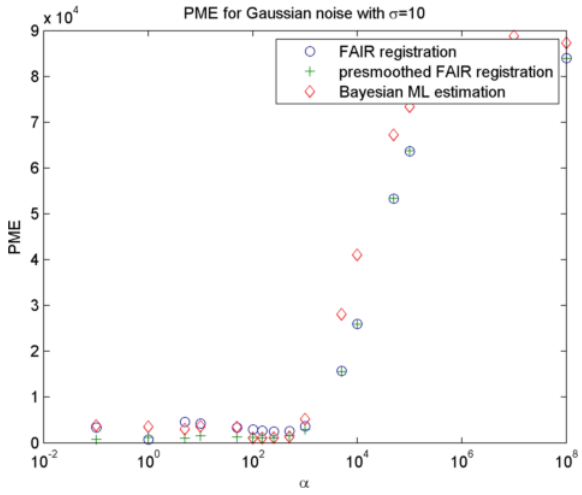
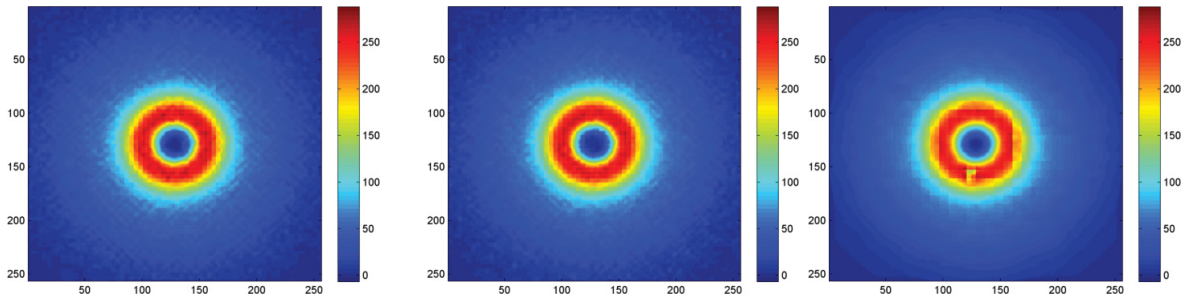


Figure 6.4: Phantom matching errors for the three registration methods and different registration parameters α.

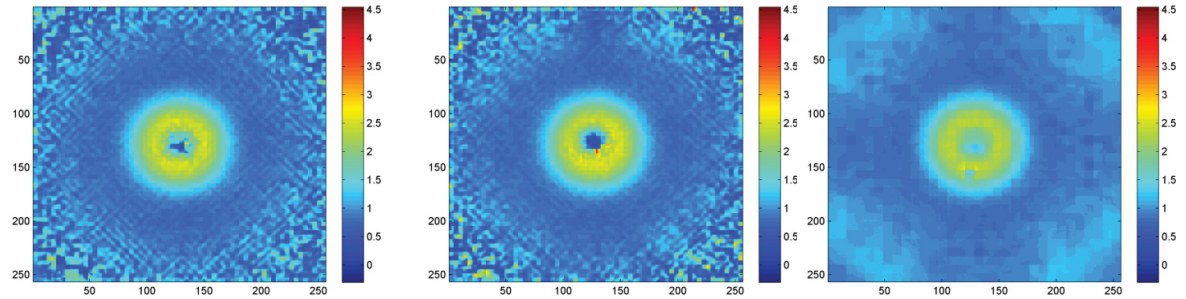
We can see that in terms of the PME all methods deliver a good registration result for a wide parameter range. In order to inspect these registration results, we show the transformed ground truth image with the lowest phantom matching error for each method:



(a) FAIR $\alpha = 1$ (b) Presmoothed FAIR $\alpha = 0.1$ (c) Bayesian ML estimation $\alpha = 100$

Figure 6.5: Ground truth template transformed with the transformation with smallest phantom matching error for each method.

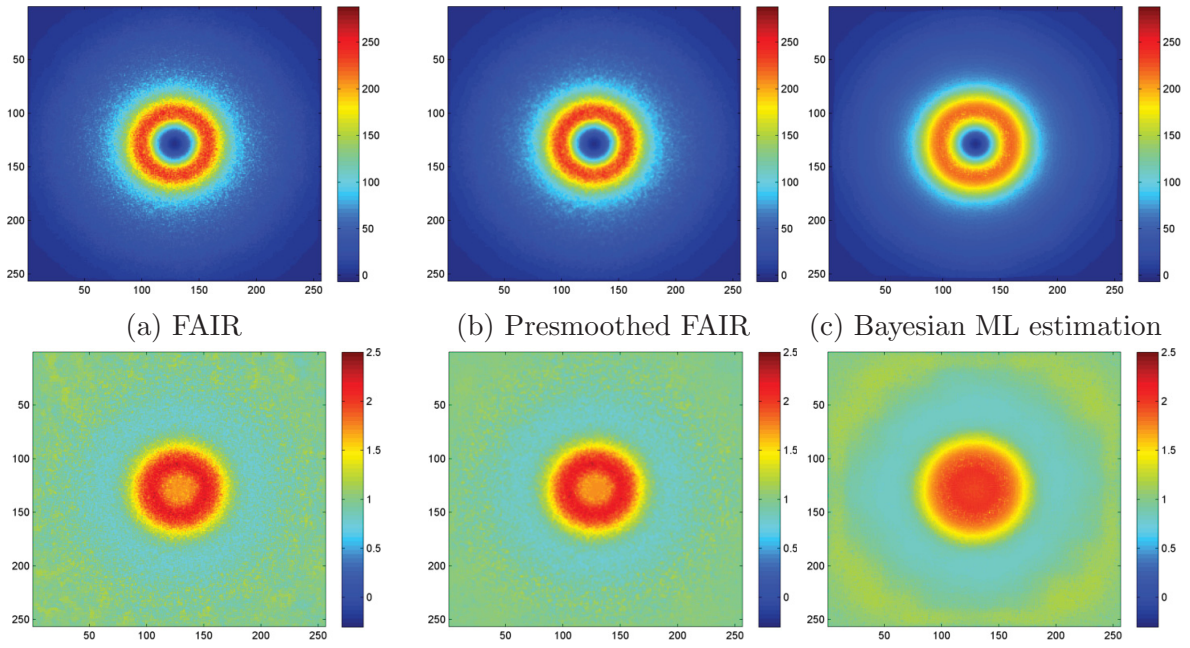
We see that the Bayesian ML estimation leads to a transformed template with a little irregularity at the bottom of the corner but sharp edges of the circle. As we see in Figure 6.6 this irregularity can be explained by the intensity modulation of the mass-preserving transformation operator:



(a) FAIR $\alpha = 1$ (b) Presmoothed FAIR $\alpha = 0.1$ (c) Bayesian ML estimation $\alpha = 100$

Figure 6.6: Jacobian determinant of the transformation with smallest phantom matching error for each method.

Note that the small irregularity at the bottom of the transformed template is observable in the Jacobian determinant as well (6.6c). Additionally the proposed method leads to a transformation with a much smoother determinant, but as we see this transformation is resulting from a registration with higher regularization weight than the other two. For a better comparison of this regularity we show transformed template and Jacobian determinant for the same regularization weight:



(a) FAIR (b) Presmoothed FAIR (c) Bayesian ML estimation

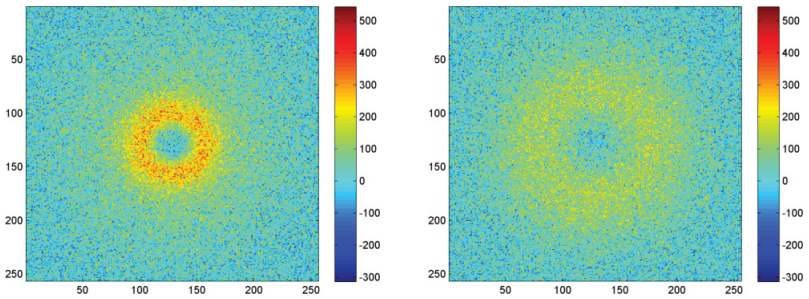
(d) FAIR (e) Presmoothed FAIR (f) Bayesian ML estimation

Figure 6.7: Transformed ground truth (top row) and Jacobian determinant for $\alpha = 1000$.

The higher regularization weight leads to smoother Jacobian determinants and as a direct consequence to a more homogeneous transformed ground truth. Note that we face a trade-off, between the regularity of the transformation, which is linked to the value of the Jacobian determinant, and a good matching of the ground truth images, which requires the Jacobian determinante to take values up to 2.5.

Average Noise

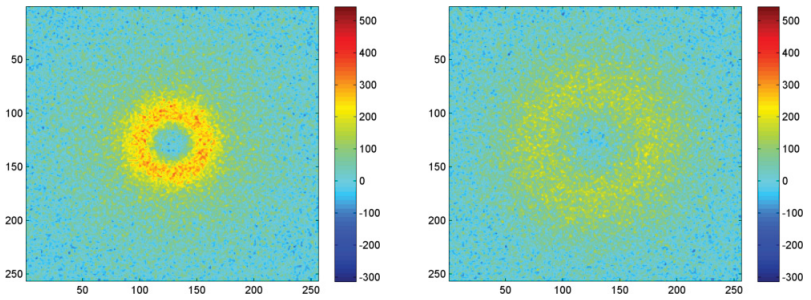
We corrupt the images now with stronger noise:



(a) Noisy reference (b) Noisy template

Figure 6.8: Noisy versions of the ground truth images to be registered. Images corrupted by Gaussian noise with standard deviation $\sigma = 80$.

While the circle in the reference image still stands out clear, it is hard to distinguish the circle from the background in the template image. Applying a Gaussian filter smoothes the images and we get:



(a) Filtered reference (b) Filtered template

Figure 6.9: Filtered versions of the noisy images in Figure 6.8.

Note that the filtering reduces the noise in both images significantly. But as a consequence of the Gaussian filtering the edges of the images get blurred, especially in the template image 6.9b.
We start again by inspecting the phantom matching errors:

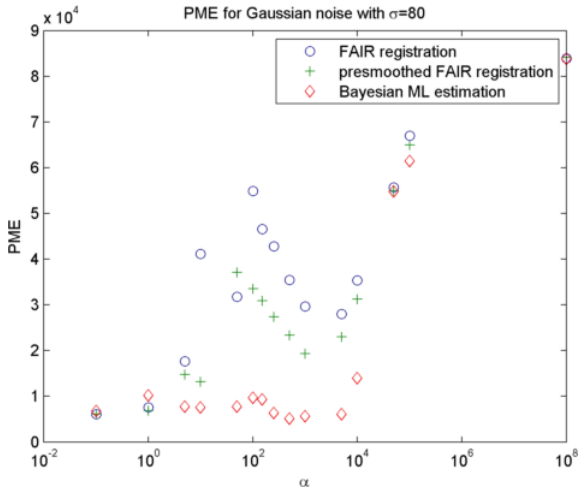


Figure 6.10: Phantom matching errors for the three registration methods and different registration parameters α.

Note that these errors are similar for small regularization parameters, but get blown up for the two classical methods with increasing α. Thus the Bayesian ML estimation leads to a more stable parameter range, where the underlying true images are nearly matched. Before we inspect the resulting transformations visually we assess the smoothness of the transformations by the regularization energy:

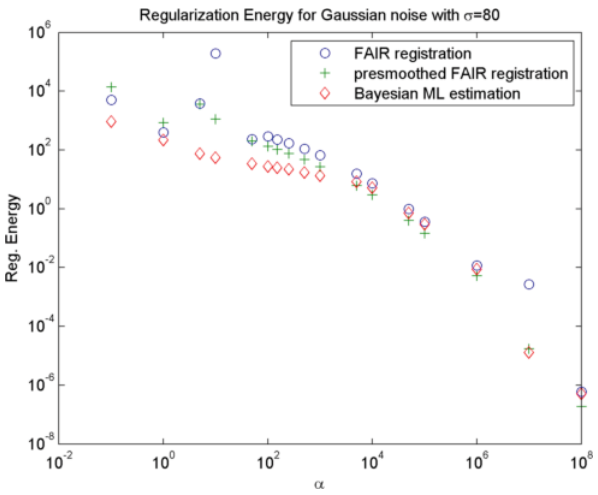
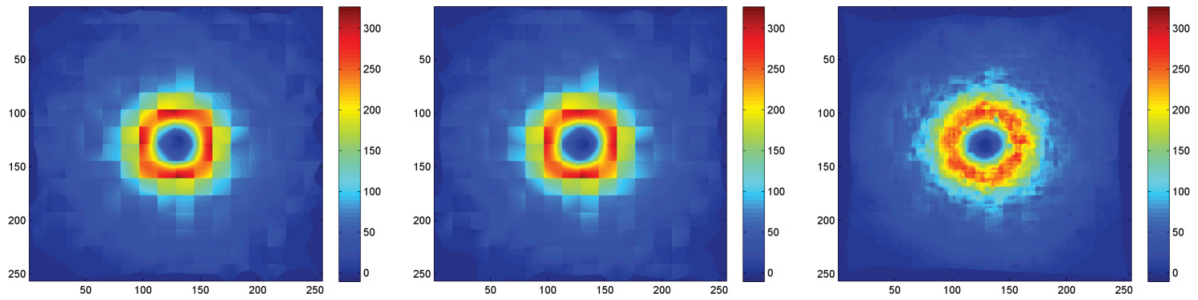


Figure 6.11: Double logarithmic plot of the regularization energy of the found transformation for each method. The regularization energy for the transformations obtained by Bayesian modeling is lower in the interval, where the phantom matching error is also smaller.

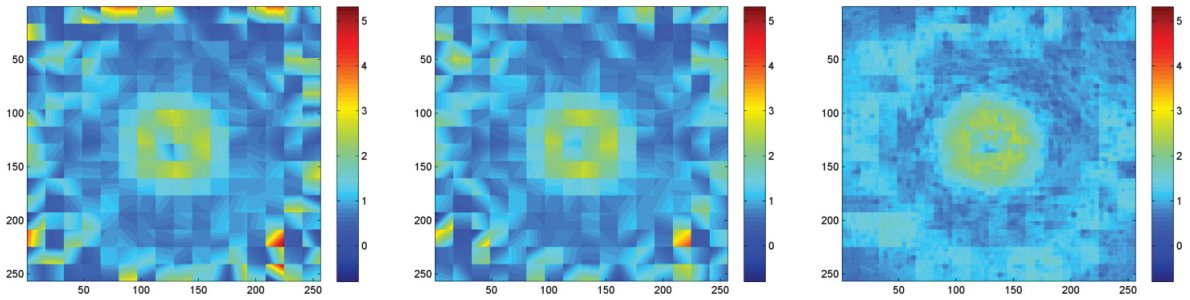
As we see the regularization energy decreases similarly for all transformation models. Note that the Bayesian ML estimation produces transformations with significantly lower regularization energy in the same interval where the phantom matching error is significantly smaller.
Figure 6.11 suggests that we can neglect the transformations, which minimize the phantom matching error for the (presmoothed) classical registration.



(a) FAIR $\alpha = 0.1$ (b) Presmoothed FAIR $\alpha = 0.1$ (c) Bayesian ML estimation $\alpha = 500$

Figure 6.12: Ground truth transformed with the transformation with smallest phantom matching error for each method. The blocky structure is a consequence of the multiplication with the Jacobian determinant.

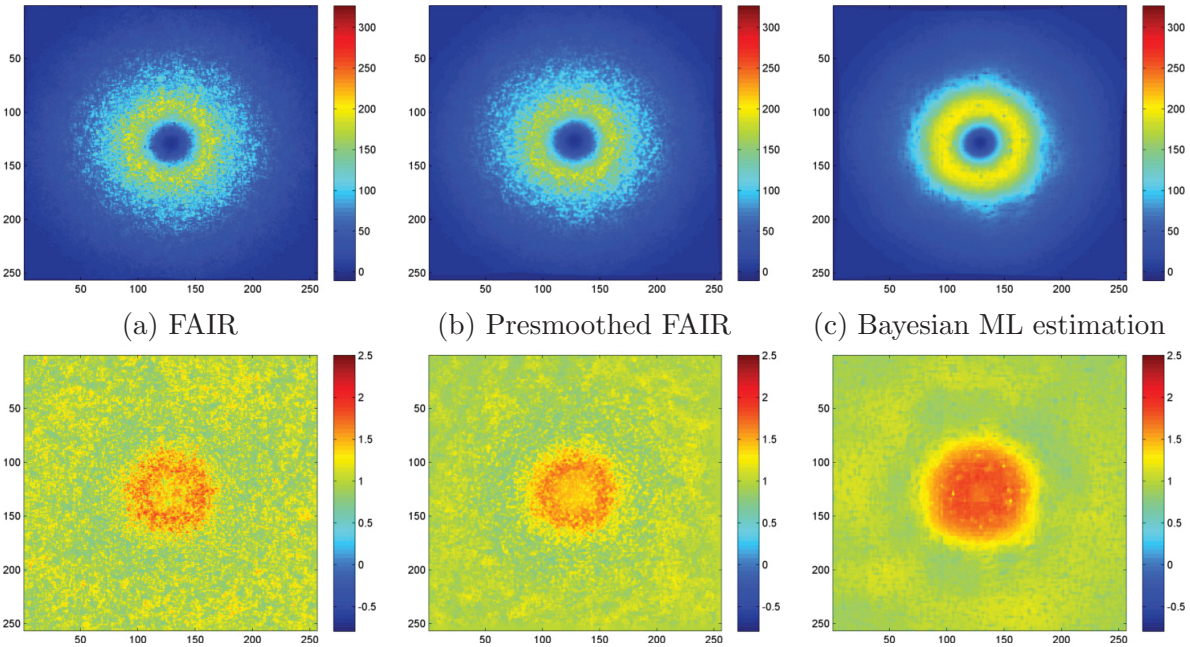
In order to explain the blocky structure of the transformed ground truth we examine the Jacobian determinant:



(a) FAIR $\alpha = 0.1$ (b) Presmoothed FAIR $\alpha = 0.1$ (c) Bayesian ML estimation $\alpha = 500$

Figure 6.13: Jacobian determinant of the transformation with smallest phantom matching error for each method.

While the determinant resulting from the proposed method is much smoother, the other two methods yield a determinant with blocky structure. This can be explained by the fact that the starting value cannot be optimized on the higher levels. Thus the algorithm was pulled to an unwanted starting value on the lower levels, which could not be improved by the optimization algorithm on the finest level.
Note that these irregularities in the Jacobian determinants could be expected, since these transformations are regularized with a small weight. Thus we visualize the estimated transformations for a bigger regularization parameter:



(a) FAIR (b) Presmoothed FAIR (c) Bayesian ML estimation

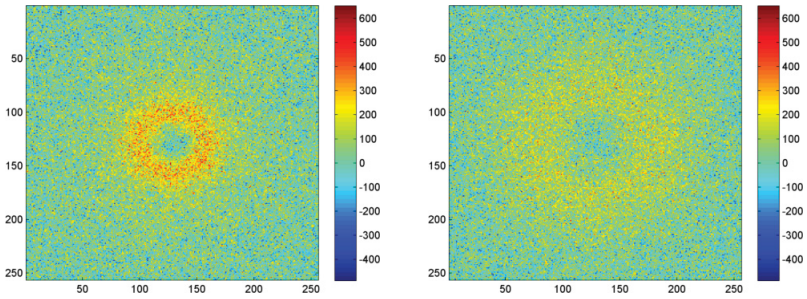
(d) FAIR (e) Presmoothed FAIR (f) Bayesian ML estimation

Figure 6.14: Transformed ground truth (top row) and Jacobian determinant for $\alpha = 10000$. The Bayesian approach results in a much smoother transformation.

Even for a high regularization weight the Bayesian approach produces much smoother Jacobian determinants. Again the matching is not exact, because the higher regularization weight punishes big Jacobian determinants too hard and thus the needed value of 2.5 is not achieved. Nevertheless only the Bayesian ML estimation preserves the contour of the reference image.

High Noise

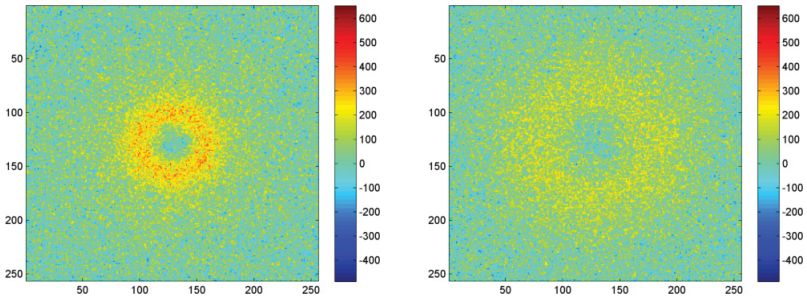
We conclude the study on the synthetic example by corrupting the images with heavy noise:



(a) Noisy reference (b) Noisy template

Figure 6.15: Noisy versions of the ground truth images to be registered. Images corrupted by Gaussian noise with standard deviation $\sigma = 125$. Especially the circle in the template image is nearly not distinguishable from the background.

Because of this heavy noise corruption the rings nearly vanish in the background. Since the intensity in the template is lower than in the reference, the reference image stays clearer. First we try to remove the noise by filtering:



(a) Filtered reference (b) Filtered template

Figure 6.16: Filtered images. The filtering improves the quality of the reference image, while the ring in the template only gets more blurred.

We see again that filtering slightly improves the sharpness of the ring in the reference image, while no visual improvement in the template image can be seen. After this preprocessing steps we turn our focus again on the comparison of all three methods.

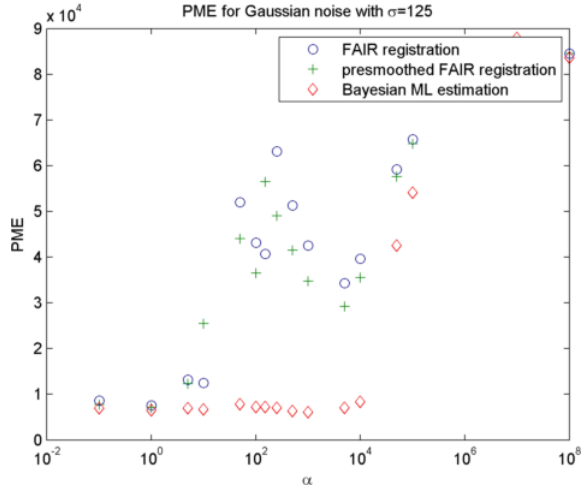


Figure 6.17: Phantom matching errors for the three registration methods and different registration parameters α.

Again we see that the classical methods only produce transformations with a significantly lower phantom matching error for very small regularization parameters. Inspecting the regularization energy suggests to neglect those transformations:

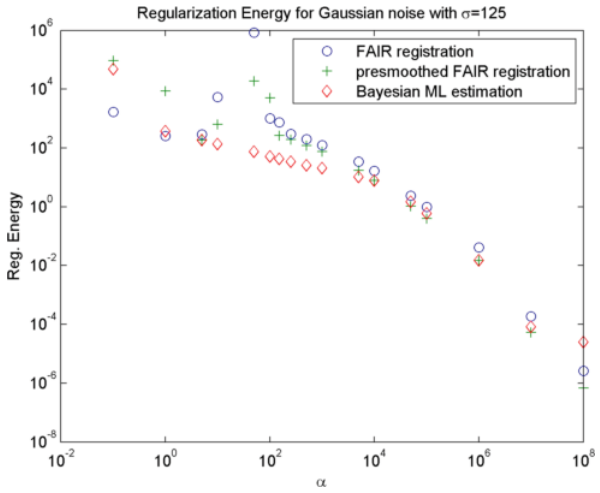
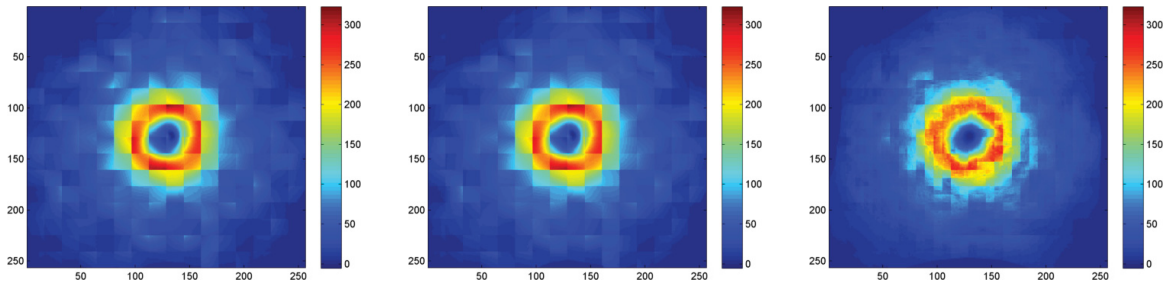


Figure 6.18: Double logarithmic plot of the regularization energy of the found transformation for each method. Note that the regularization energy for the Bayesian ML estimation is significantly lower than the regularization energy for the other two methods for $\alpha < 10^4$.

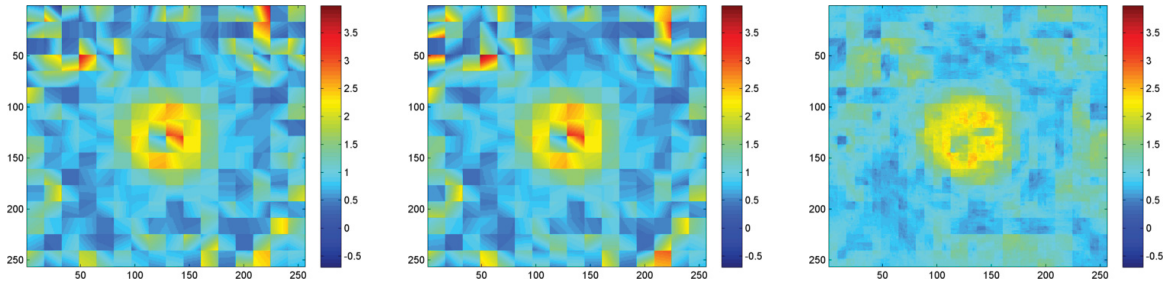
We see that the regularization energy of the transformations resulting from the Bayesian ML estimation is significantly lower than the regularization energies for the other transformations resulting from the other two methods. To give a more visual impression we show the transformed ground truth images with the transformation minimizing the phantom matching error for each registration method:



(a) FAIR $\alpha = 1$ (b) Presmoothed FAIR $\alpha = 1$ (c) Bayesian ML estimation $\alpha = 1000$

Figure 6.19: Ground truth transformed with the transformation with smallest phantom matching error for each method. The blocky structure is a consequence of the multiplication with the Jacobian determinant.

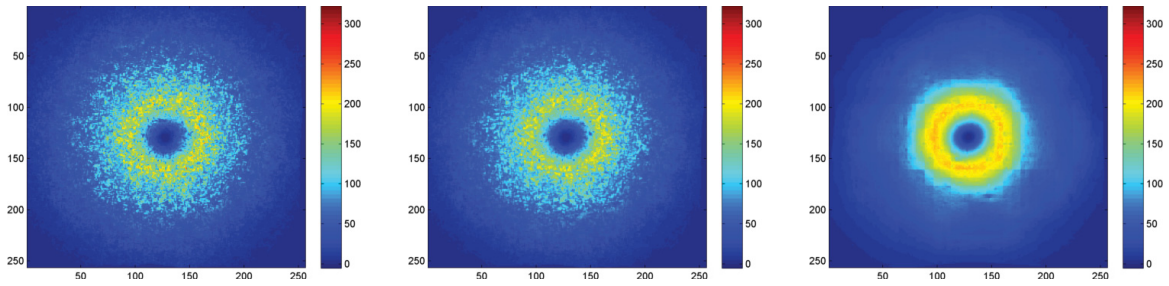
Again the transformed ground truth image suffers from a blocky structure. In order to explain this structure, we proceed in the same way as before and inspect the Jacobian determinants for all transformations:



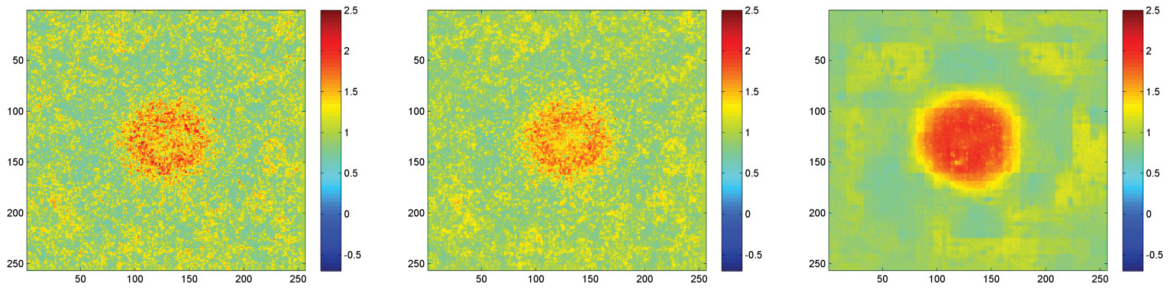
(a) FAIR $\alpha = 1$ (b) Presmoothed FAIR $\alpha = 1$ (c) Bayesian ML estimation $\alpha = 1000$

Figure 6.20: Jacobian determinant of the transformation with smallest phantom matching error for each method. Note the blocky structure of the Jacobian determinant for the two classical approaches.

Since the regularization parameter for the Bayesian ML estimation is significantly higher we compare the three approaches also for a fixed regularization parameter:



(a) FAIR (b) Presmoothed FAIR (c) Bayesian ML estimation



(d) FAIR (e) Presmoothed FAIR (f) Bayesian ML estimation

Figure 6.21: Transformed ground truth (top row) and Jacobian determinant for $\alpha = 10000$. The Bayesian approach results in a much smoother transformation.

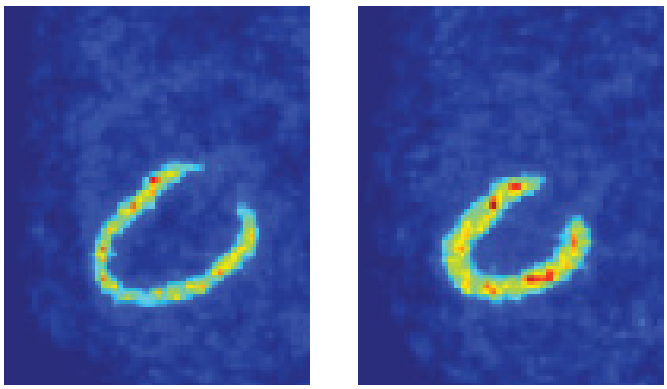
Again we see that even for high regularization weights the Bayesian approach leads to a much smoother Jacobian determinant and therefore preserves the contour of the circle. Thus the registration with noise modeling gives stable registration results for a wide parameter range.

Concluding Assessment

All three methods produced similar results for a very low noise level. Thus in this case all three methods are a suitable choice.
By driving the noise level higher we were able to observe different results for the three methods. For the classical methods the transformations best in the sense of the phantom matching error were produced by registrations with small regularization parameter. Consequently the noise structure could be observed in a irregular Jacobian determinant. But even stronger regularization could only mildly alleviate this. In contrast to that Bayesian ML estimation produced even in the case of high noisy corruption smooth transformations for a wide parameter range (see Figures 6.11 and 6.18). This can be explained with the fact, that implicit image estimation in the proposed method manages to match the underlying images for the low regularization weights, while the classical methods are still trying to match the noise.
A remaining difficulty is to balance the regularization weight, such that the determinant takes values in the desired range for an exact matching of the intensity and not only the contour.

6.5.2 XCAT Software Phantom

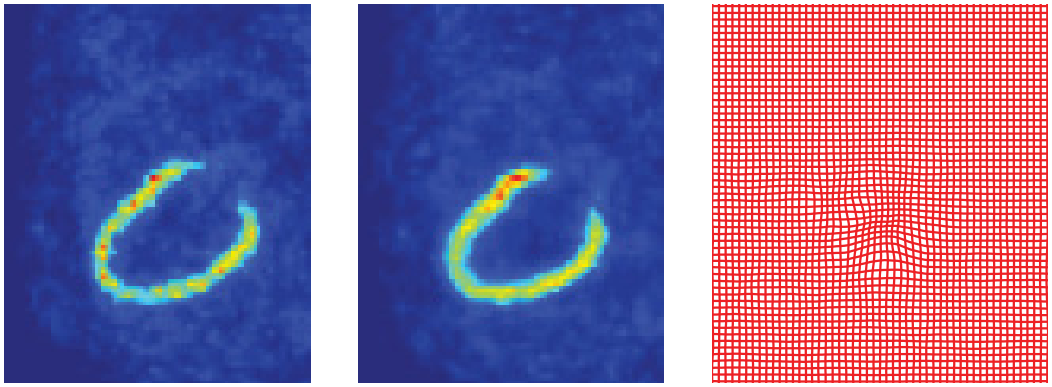
As a proof of concept we again shortly present the results on the XCAT phantom [134] we already presented in [142]. We generated simulated PET images of the end-diastolic phase (Reference, 6.22a) and the end systolic-phase (Template, 6.22b).



(a) Reference (b) Template

Figure 6.22: Representative 2D slices of PET images, generated by the XCAT software phantom [134]. While the end diastolic phase serves as reference image, the end systolic phase gets transformed.

Both images have the data size $175 \times 175 \times 47$, thus the images in Figure 6.23 are only representative 2D slices. Applying the ML estimation leads to the following deformation grid and transformed template:



(a) Ground truth reference (b) Transformed template (c) Transformation grid

Figure 6.23: Representative 2D slices of the transformed template (left) and corresponding transformation grid (right). As the slices indicate, the transformation is smooth and maps the template nearly exactly to the reference.

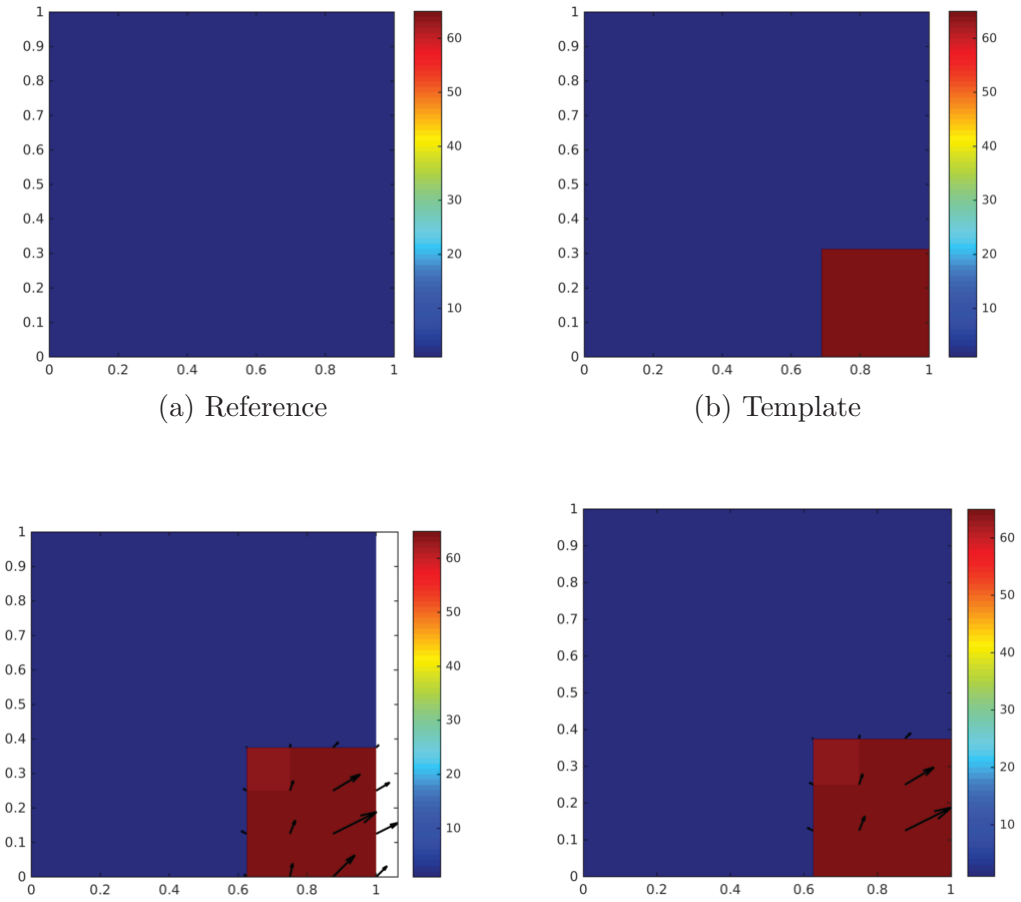
We see that the Bayesian ML estimation generates a smooth transformation, which maps the template nearly exactly to the reference image. Thus we gave a proof of concept for the proposed method; a thorough evaluation of the method on real data has to be carried out in future studies.

6.6 Discussion and Outlook

We presented an unified framework for MAP-estimation in image registration with noise modeling. After computing registration functionals for given data fidelities for some specific noise models, we turned our focus to the analysis of the resulting variational problem. We were able to show that we can expect the existence of a minimizer for suitable data fidelities, if we use a mass-preserving transformation model. The analytical results for the intensity-constant transformation model are raising more questions: Since we proved that the resulting distance measures are concave with respect to the Jacobian determinant for convex data fidelities, it is much more complicated to give existence results for this case. After presenting a numerical framework we evaluated our methods on a synthetic toy example and motivated the use of this framework for real data.
While similar ideas have been presented earlier in the literature, our main contribution lies in the unification for different noise models and the analytical results regarding the polyconvexity of the registration functionals. A coherent statistical framework for Gaussian noise has been presented by [3]. While in [123] incorporation of noise characteristics into distance measures has been investigated, several registration criteria for parametric transformations have been presented in [135]. Note that the latter two approaches do not impose regularization on the motion.
There are several open problems concerning the registration method we described in this chapter. In the following we will shortly discuss some of these and hint at possible further research in this field.

6.6.1 Injectivity in Image Registration

The framework we presented so far is neglecting the assumption, that the transformation is injective. This assumption is crucial, since we assumed injectivity in the framework we presented in this chapter. However a restriction to injective transformations is difficult to realize. One possibility is to restrict ourselves to a minimization problem with injective boundary values (compare Corollary 7.2.7). Then the injectivity in the interior is given by Proposition 3.2.9.
The question now arising is how to determine these boundary values in practical applications. If we observe a larger image than the region of interest, where the motion occurs, we can restrict the registration to transformations v with $v|_{\partial\Omega} = Id$. If this is not the case, performing an affine linear preregistration with regularization of the Jacobian determinant is an option: The regularization of the determinant grants locally invertible affine linear transformations, which are directly global invertible affine linear transformations. The result of the affine linear preregistration can now serve as an injective boundary value.
A straightforward way to incorporate this into the numerical framework we presented is to set the BFGS update to zero at the boundary. However this might cause problems: If we have an object close to the boundary and the force field generating the search direction in the distance term pushes this object outside of the image domain.

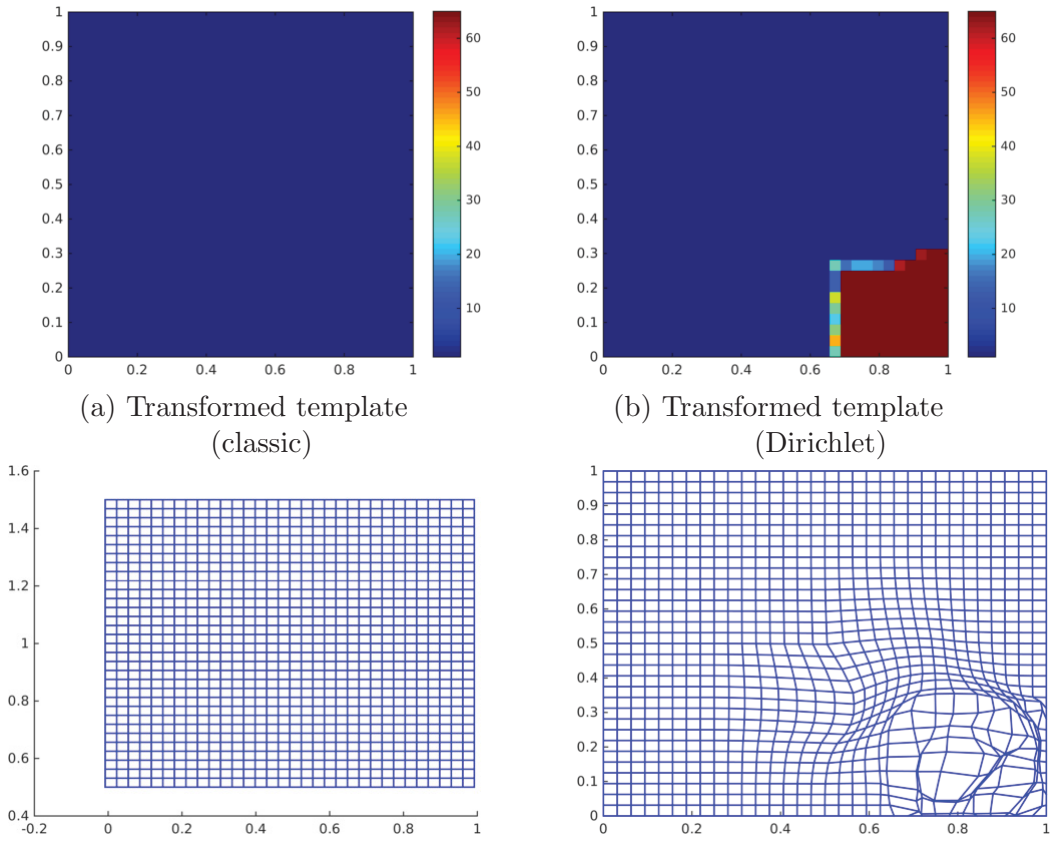


(a) Reference (b) Template

(c) Search direction (d) Search direction (Dirichlet)

Figure 6.24: Artificial image registration problem (top row): Since the reference image is constant zero, we expect the square in the template to be pushed outside of the image domain. The initial search direction for the normal (left) approach and Dirichlet boundary conditions (right) is visualized by black arrows (bottom row).

As we see in Figure 6.24 there is only a non-vanishing search direction inside small the square. Note that the search direction points outwards of the image domain in both cases. While in the classic approach the update vector on the boundary points outside of the image domain (white strip to the right in 6.24c), this update vector is set to zero in the approach with Dirichlet boundary conditions. Consequently we expect severe problems for the Dirichlet registration, since the update vector pulls the grid towards foldings, which are excluded by the hyperelastic regularizer:



(a) Transformed template (classic)

(b) Transformed template (Dirichlet)

(c) Transformation grid

(d) Transformation grid (Dirichlet)

Figure 6.25: Registration results (top row) with transformation grids (bottom row) for classic and Dirichlet approach with $\alpha_1 = \alpha_3 = 10$ and $\alpha_2 = 0$. The classic method (left) pushes the square outside of the domain, while the border cannot be crossed without violation of the determinant constraint ($\det(\nabla v) > 0$) in the Dirichlet approach.

Note that the irregularities in the grid 6.25d can be explained in the following way: The forces pulling the square outside are stronger in the middle of the center, which results in enhancement of the corresponding grid cells. Consequently the grid cells near the boundary get shrinked and moved towards the boundary. Motivated by this we observe that the approach outlined above may be obstructed by the following problems:

- How can we deal with forces trying to push an object outside of the domain, while the boundary is fixed by the Dirichlet boundary values?
- What can we do, when no injective Dirichlet boundary values are available?

These problems give rise to the question, whether the injectivity of a starting guess is preserved in the optimization process. As we observe the defined transformation operators T^{ic} and T^{mp} for a transformation v are injective, iff v is weakly invertible. Furthermore as a consequence of the Neumann series [146, Chapter II, Theorem 1.11] the set

of invertible operators is open with respect to the operator norm topology. Additionally for a given injective operator T_1, every operator T_2 with $\|T_1 - T_2\| \leq q\|(T_1)^{-1}\|^{-1}$ for $0 < q < 1$ is injective and we have:

$$\|(T_2)^{-1}\| \leq \frac{1}{1-q}\|(T_1)^{-1}\|^{-1} \tag{6.40}$$

This estimation above might prove to be too restrictive to be used to derive injectivity conditions in our numerical framework, since it does not take the local invertibility of the admissible transformations into account. Thus the question arises, if we can derive conditions for an additive update s, such that $v_1 + s$ is an admissible transformation and stays injective for an injective transformation v_1. On the other hand one could try to find a registration problem generating an update s for an injective transformation v, such that $(v + s) \in \mathscr{A}$, but $v + s$ is not injective.
To complete this short outlook on injectivity in image registration, we mention that there is a possibility to guarantee injectivity, but at the price of increasing the computational costs: We can embed the images to be registered into bigger images with a constant background. Then we can safely assume that the transformation is the identity on the boundary of the bigger images and Proposition 3.2.9 guarantees the injectivity of an admissible transformation. Additionally, this embedding can help to alleviate the difficulties induced by forces pushing a part of the image outside of the domain, since the constant background ensures, that the forces for the distance measure are zero near the boundary of the bigger image.

6.6.2 Convergence Properties

There are several questions arising regarding the convergence properties in a numerical framework. As we saw in Chapter 5 the convergence of the modified BFGS method we use in the optimization relies on the convergence of the classical BFGS method. Convergence to a local minimum for the classical method is given for an initial starting point with bounded sublevel set, provided the objective function is continuously differentiable with Lipschitz continuous gradients [91].
We divide the discussion on the convergence properties into two parts: While we discuss convergence of the modified BFGS in the first part, we address the question to which local minimum the method converges in the second.

Convergence of the Modified BFGS-Method

As we mentioned above there are three properties of the objective function j, which ensure convergence to a local minimum. According to [91] these properties are given by:

1. The objective function j is continuously differentiable.
2. ∇j is Lipschitz-continuous.
3. The sublevel set $\{x \,|\, j(x) \leq j(x_0)\}$ is bounded for the starting value x_0.

We will discuss these three conditions shortly in the following:

Condition 1 should not be a problem in practical applications: Since we start with digital images we can obtain continuous differentiability via appropriate interpolation (e.g. cubic or spline interpolation - see [100, Chapter 6.1] for a similar argument). Note that we assumed data fidelities to be continuously differentiable and thus we end up with the composition of two C^1 functions.

The second condition might cause problems for some given noise models: The Kullback-Leibler data fidelity D^{KL} for data f corrupted by Poisson noise is given by

$$D^{KL} : L^1(\Omega, \mathbb{R}_{>0}) \to \mathbb{R} \qquad u \to \int_\Omega u(x) - f(x)\log(u(x))\mathrm{d}x. \tag{6.41}$$

As we observe this data fidelity explodes for images $u \to 0$ and therefore we cannot expect Lipschitz continuity. In practical applications often a threshold ϵ is set, such that the integrand function is given by:

$$u \to \begin{cases} u(x) - f(x)\log(u(x)) & u(x) > \epsilon \\ 0 & u(x) \leq \epsilon \end{cases}. \tag{6.42}$$

As we directly observe this function is not even differentiable for $u(x) = \epsilon$, but we can achieve a Lipschitz-continuous approximation by choosing two thresholds $\epsilon_1 > \epsilon_2$, setting the integrand function to zero for $u(x) \leq \epsilon_2$ and performing an appropriate interpolation in the interval $[\epsilon_2, \epsilon_1]$.

Condition 3 can be critical in practical applications: As we see a global translation applied to all grid points has zero regularization energy. Thus we can find arbitrarily many transformation grids with a finite distance term and zero regularization energy by simply performing a translation, which pushes the original image outside of the image domain. As a short remark we present two ways to handle these difficulties: The first one is to alternate boundary conditions for the interpolation, which transforms the image in a manner, such that the values of the transformed image are set to infinity, if the grid points are pushed too far from the original image domain. Another possibility is to focus on the regularization energy of said translations. Ruthotto [129] proposed to restrict the admissible transformations to transformations with bounded displacements (compare (4.68) in the definition of the hyperelastic regularization) by requiring

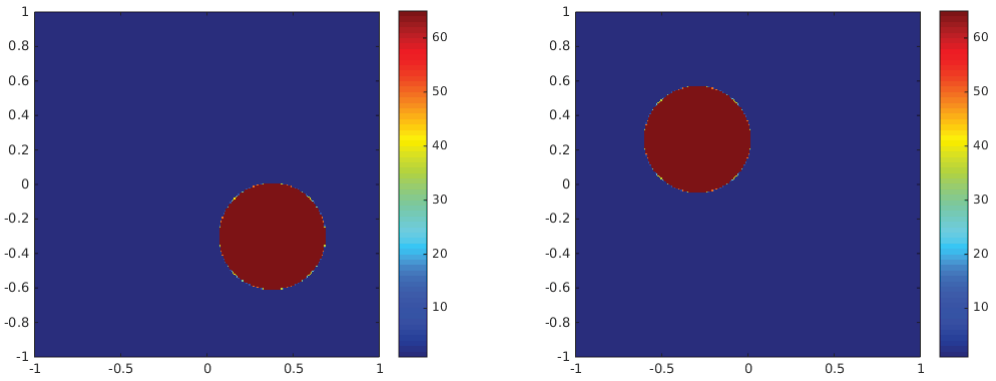
$$\int_\Omega y(x)dx| \leq \mathrm{vol}(\Omega)(M + \mathrm{diam}(\Omega)).$$

It is clear that this restriction discards the unwanted translations we described earlier.

We have now shown that the minimization method we use in our framework converges to a local minimum, given the assumptions we outlined above. Note that also for a line search with Wolfe conditions [91] as well as for nonmonotone Armijo line search [149] global convergence to a local minimum can be shown. The next question is, which local minimum is chosen by the minimization method.

A finer Multilevel Sequence

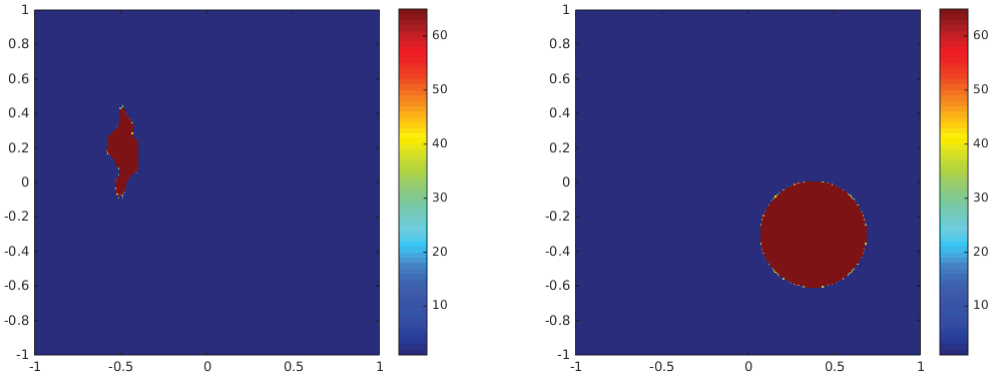
As we described in Chapter 5 we perform a multilevel approach in the minimization process, because we hope to decrease the probability of obtaining unwanted local minima. The following example shows that the multilevel approach is even sensitive to the scaling factor s between two levels:



(a) Reference (b) Template

Figure 6.26: Registration problem to compare the impact of different scaling factors in the multilevel framework.

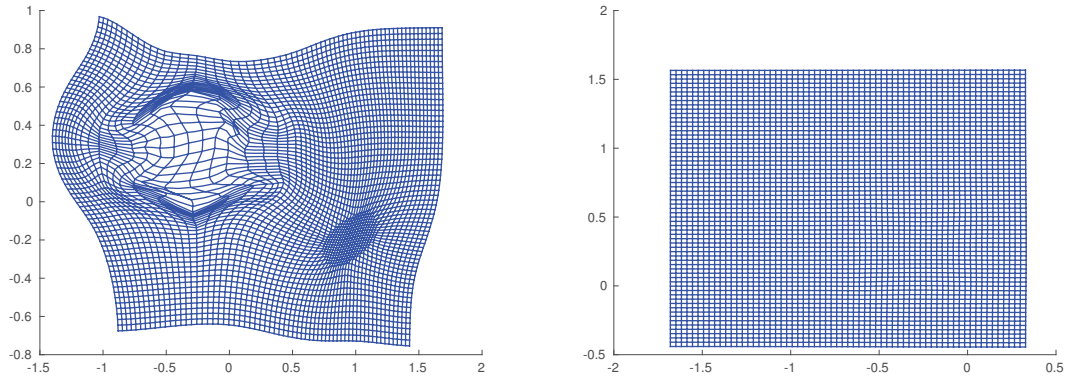
We perform a multilevel registration with $\alpha = 100$ and different scaling factors s between the levels:



(a) Transformed template ($s = 2$) (b) Transformed template ($s = 1.2$)

Figure 6.27: Transformed template images (top row) with transformation grids (bottom row) for different scaling factors s. The finer sequence reduces the objective function to 0.037, while the coarse sequence produces a total energy of 4726.

Before we compare both approaches we inspect the transformation grids of the found transformation.



(a) Transformation Grid ($s = 2$)

(b) Transformation Grid ($s = 1.2$)

Figure 6.28: Transformation grids for both multilevel sequences. While the finer sequence leads to a translation, the coarser sequence produces an irregular transformation.

We see that for the coarser scaling $s = 2$ the algorithm tries to minimize the distance by compressing the circle nearly into a small dot. In contrast to that a finer scaling in the multilevel frameworks generates a transformation aligning the template nearly perfectly to the reference image. Note that a similar result can be produced with the coarse framework by affine linear registration. However, we can state, that a finer multilevel sequence can rule out unwanted local minimizers, but of course at higher computational costs. Motivated by this synthetic example the question arises, whether this differences between coarse and fine multilevel scaling can influence registration results in practical applications.

6.6.3 Noise and Multilevel Approaches

We performed our multilevel minimization with an adaptive parameter choice, where we linked the regularization parameter α to the variance of the noisy data given on each level. We motivated this adaptive choice for data corrupted by Gaussian noise with Example 5.1.1. Unfortunately, for other noise types it is not straightforward how the variance behaves in the multilevel framework. If we consider an image $f = f_{i,j}$ corrupted by Poisson noise, then the corruption is given by

$$f_{i,j} = \mathrm{Poi}(u_{i,j}) \tag{6.43}$$

with $u = u_{i,j}$ being the underlying true image. Thus the random variables generating the noise in each pixel, are not identically distributed, since the mean and variance are given by the value of the true image. While the behaviour of Poisson noise in multilevel approaches has been studied for reconstruction problems [52, 33], to our best knowledge this has not been investigated for registration problems yet.

Another aspect of the noise modeling in the multilevel approach is, that it is not always clear, how the noise characteristics change, if we obtain a coarser image by averaging:

While the sum of Poisson distributed random variables stays Poisson distributed, this property can not be expected to be preserved by averaging. One method to deal with this difficulties is applying the Anscombe transformation (compare e.g. [52]), which transformes Poisson distributed random variables in normal distributed ones. The drawback of the Anscombe transformation is that the variance is increased by the transformation. Since we formulated a unified framework for registration with noise modeling, the question arises how the noise characteristics change by the averaging resulting from the multilevel approach. Directly arising from this is the question, how the change of the noise characteristics affects the data fidelities we used for a specific noise model.

As we outlined shortly the multilevel approach needs to be studied in two aspects: The first one is how the variance of the noisy signal changes by the averaging in the multilevel approach. While this is understood for Gaussian noise, research for other noise models needs to be done. The second aspect is closely related: Do we need to use specific data fidelities for each level, since the noise characteristics change on each level?

6.6.4 Adjoint Operators for Non-Injective Transformations

In the derivation of the registration formulation for noisy image sequences we assumed, that the transformation we estimate is one-to-one. This assumption had the advantage, that an inverse existed and thus the computation of the adjoint operator with the area formula was straightforward. In the following we outline a way to compute adjoints to transformation operators for non-injective transformations.

For a not necessarily injective transformation v we assume, that we can decompose the image domain Ω into disjoint sets, such that

$$\Omega = \bigcup_{i}^{n} \Omega_i \cup N \tag{6.44}$$

with $\lambda(N) = 0$ and $v|_{\Omega_i} =: v_i$ injective. Then we can start to compute the adjoint operator for the intensity-constant transformation operator via

$$\begin{aligned}
\int_{\Omega} f(v(x))g(x)\mathrm{d}x &= \int_{\bigcup_i^n \Omega_i \cup N} f(v(x))g(x)\mathrm{d}x \\
&= \int_{\bigcup_i^n \Omega_i} f(v(x))g(x)\mathrm{d}x \\
&= \sum_{i}^{n} \int_{\Omega_i} f(v_i(x))g(x)\mathrm{d}x
\end{aligned}$$

$$
\begin{aligned}
&= \sum_i^n \int_{v_i(\Omega_i)} f(x) g(v_i^{-1}(x)) \det(\nabla v_i^{-1}) \mathrm{d}x \\
&= \sum_i^n \int_{v_i(\Omega_i)} f(x) g(v_i^{-1}(x)) \frac{1}{\det(\nabla v_i)} \mathrm{d}x \\
&= \int_{v(\Omega)} f(x) \left(\sum_{z \in v^{-1}(x)} g(z) \right) \frac{1}{\det(\nabla v)} \mathrm{d}x. \qquad (6.45)
\end{aligned}
$$

For computing the adjoint in the case of the mass-preserving transformation operator we can proceed analogously and obtain

$$
\int_\Omega f(v(x)) g(x) \det(\nabla v) \mathrm{d}x = \int_{v(\Omega)} f(x) \left(\sum_{z \in v^{-1}(x)} g(z) \right) \mathrm{d}x. \qquad (6.46)
$$

Note that the decomposition (6.44) cannot be found, if $N_v(\cdot, \Omega) \notin L^\infty(\Omega)$. Thus there remain two open questions to be studied:

1. Can we find a decomposition (6.44), if $N_v(\cdot, \Omega) \in L^\infty(\Omega)$?
2. What can we do, if $N_v(\cdot, \Omega) \notin L^\infty(\Omega)$?

Since the boundedness of the Banach indicatrix in L^∞ is sufficient for the transformation operator to be continuous we will not pursue the second question further, but rather focus on the first. We start by stating the aforementioned continuity property of the mass-preserving transformation operator.

Lemma 6.6.1. *Let $v : \Omega \to \hat{\Omega}$ be an admissible transformation with $N_v(\cdot, \hat{\Omega}) \in L^\infty$ as well as $\det(\nabla v) \in L^\infty$. Then the mass-preserving transformation operator*

$$
T_v^{mp} : L^2(\hat{\Omega}) \to L^2(\Omega) \qquad (6.47)
$$

is continuous.

Proof. We observe:

$$
\begin{aligned}
\|T_v^{mp} u\|_{L^2(\Omega)}^2 &= \int_\Omega (T_v^{mp} u)^2 \, \mathrm{d}x \\
&= \int_\Omega \left(u(v(x)) \det(\nabla v(x)) \right)^2 \mathrm{d}x \\
&\leq \| \det(\nabla v) \|_\infty \int_\Omega \left(u(v(x)) \right)^2 \det(\nabla v) \mathrm{d}x
\end{aligned}
$$

$$
\begin{aligned}
&= \|\det(\nabla v)\|_\infty \int_{\mathbb{R}^d} (u(z))^2 \, N_v(z, \hat{\Omega}) \mathrm{d}z \\
&\leq \|\det(\nabla v)\|_\infty \|N_v(\cdot, \hat{\Omega})\|_\infty \int_{\mathbb{R}^d} u(z)^2 \mathrm{d}z \\
&= \|\det(\nabla v)\|_\infty \|N_v(\cdot, \hat{\Omega})\|_\infty \|u\|^2_{L^2(\hat{\Omega})}.
\end{aligned}
$$

□

This continuity property ensures that it is sufficient to verify the adjoint operator (6.46) for step functions g. Thus for any step function $g = \sum_{k=1}^{K} c_k \mathcal{X}_{E_k}$ we obtain:

$$
\begin{aligned}
\int_\Omega u(v(x)) \det(\nabla v(x)) g(x) \mathrm{d}x &= \sum_{k=1}^{N} c_k \int_{E_k} u(v(x)) \det(\nabla v(x)) \mathrm{d}x \\
&= \sum_{k=1}^{N} c_k \int_{\mathbb{R}^d} u(z) N_v(z, E_k) \mathrm{d}z \\
&= \int_{\mathbb{R}^d} u(z) \sum_{x \in v^{-1}(z)} u(x) \mathrm{d}z.
\end{aligned}
$$

We can proceed analogously for the intensity-constant operator, if we assume $\det(\nabla v)^{-1} \in L^\infty$. To conclude this discussion we will compute the resulting distance measure for a not necessarily injective transformation in the case of Gaussian noise and the mass-preserving transformation operator.
We compute analogously to Section 6.2 and start by inserting the modified transformation operators in the optimality condition (6.16) for Gaussian noise:

$$
\begin{aligned}
0 &= 2(\overline{u}(x) - f_0(x)) + 2 \left(T_v^{mp}\right)^* \left(\overline{u}(v(x)) \det(\nabla v(x)) - f_1(x)\right) \\
0 &= \overline{u}(x) - f_0(x) + \overline{u}(x) \frac{N_v(x, \Omega)}{\det(\nabla v^{-1}(x)} - \sum_{z \in v^{-1}(x)} f_1(z) \\
\left(1 + \frac{N_v(x, \Omega)}{\det(\nabla v^{-1}(x))}\right) \overline{u} &= f_0(x) + \sum_{z \in v^{-1}(x)} f_1(z) \\
\overline{u} &= \frac{f_0(x) + \sum_{z \in v^{-1}(x)} f_1(z)}{1 + \frac{N_v(x, \Omega)}{\det(\nabla v^{-1}(x))}} \\
\overline{u} &= \frac{f_0(x) \det(\nabla v^{-1}(x)) + \sum_{z \in v^{-1}(x)} f_1(z) \det(\nabla v^{-1}(x))}{N_v(x, \Omega) + \det(\nabla v^{-1}(x))}.
\end{aligned}
$$

We start by computing the transformed version of $\overline{u}$:

$$\begin{aligned}\overline{u}(v(x)) &= \frac{f_0(v(x))\frac{1}{\det(\nabla v(x))} + f_1(x)\frac{N_v(v(x),\Omega)}{\det(\nabla v)}}{N_v(v(x), v(\Omega)) + \frac{1}{\det(\nabla v)}} \\ &= \frac{f_0(v(x)) + f_1(x)N_v(v(x),\Omega)}{N_v(v(x), v(\Omega))\det(\nabla v) + 1}.\end{aligned}$$

We can use this transformed version to expand the first distance term:

$$\begin{aligned}D(\overline{u}, f_0) &= \int_{\hat{\Omega}} (\overline{u}(x) - f_0(x))^2 \mathrm{d}x \\ &= \int_{\mathbb{R}^d \cap \{N_v(x,\Omega)\neq 0\}} (\overline{u}(x) - f_0(x))^2 \frac{N_v(x,\Omega)}{N_v(x,\Omega)} \mathrm{d}x \\ &= \int_{\Omega} (\overline{u}(v(x)) - f_0(x))^2 \frac{\det(\nabla v(x))}{N_v(v(x), v(\Omega))} \mathrm{d}x \\ &= \int_{\Omega} \left(\frac{f_0(v(x)) + f_1(x)N_v(v(x),\Omega)}{N_v(v(x),\Omega)\det(\nabla v) + 1} - f_0(v(x)) \right)^2 \frac{\det(\nabla v(x))}{N_v(v(x),\Omega)} \mathrm{d}x \\ &= \int_{\Omega} \left(\frac{f_1(x)N_v(v(x),\Omega) - f_0(v(x))\det(\nabla v(x))N_v(v(x),\Omega)}{N_v(v(x),\Omega)\det(\nabla v) + 1} \right)^2 \frac{\det(\nabla v(x))}{N_v(v(x),\Omega)} \mathrm{d}x \\ &= \int_{\Omega} (f_0(v(x))\det(\nabla v(x)) - f_1(x))^2 \frac{N_v(v(x),\Omega)^2 \det(\nabla v(x))}{(N_v(v(x),\Omega)\det(\nabla v(x)) + 1)^2 N_v(v(x),\Omega)} \mathrm{d}x \\ &= \int_{\Omega} (f_0(v(x))\det(\nabla v(x)) - f_1(x))^2 \frac{N_v(v(x),\Omega)\det(\nabla v(x))}{(N_v(v(x),\Omega)\det(\nabla v(x)) + 1)^2} \mathrm{d}x\end{aligned}$$

Now we evaluate the second distance term:

$$\begin{aligned}D(\overline{u}(v)\det(\nabla v), f_1) &= \int_{\Omega} (\overline{u}(v)\det(\nabla v) - f_1)^2 \mathrm{d}x \\ &= \int_{\Omega} \left(\frac{f_0(v(x)) + f_1(x)N_v(v(x),\Omega)}{N_v(v(x), v(\Omega))\det(\nabla v) + 1} \det(\nabla v(x)) - f_1(x) \right)^2 \mathrm{d}x \\ &= \int_{\Omega} \left(\frac{f_0(v(x))\det(\nabla v(x)) - f_1(x)}{N_v(v(x), v(\Omega))\det(\nabla v) + 1} \right)^2 \mathrm{d}x\end{aligned}$$

$$= \int_\Omega \frac{(f_0(v(x))\det(\nabla v(x)) - f_1(x))^2}{(N_v(v(x),\Omega)\det(\nabla v(x))+1)^2}\mathrm{d}x.$$

By adding up this two distance terms one power in the denominator cancels out and we obtain the following modified distance measure

$$\tilde{D}^{gaussian,mp} = \int_\Omega \frac{(f_0(v(x))\det(\nabla v(x)) - f_1(x))^2}{N_v(v(x),\Omega)\det(\nabla v(x))+1}\mathrm{d}x. \tag{6.48}$$

Note that the Jacobian determinant $\det(\nabla v^{-1})$ can be computed for each $x \in v^{-1}(z)$ with the help of the local inverse. As a consequence of the implicit function theorem we obtain, that the Jacobian determinants coincide for all those x. Thus we just used the notation $x \in v^{-1}(z)$ in order to keep the computations simple.
As a numerical implementation of the Banach indicatrix N_v is a challenging task, we will not elaborate further on a possible numerical framework. Nevertheless modified distance measures can be computed for the intensity-constant transformation operator and other noise models as well.
To conclude this discussion we mention that the generalized distance measure coincides for injective transformations with the one we deduced with the injectivity constraint in Section 6.2.

6.6.5 Intensity-Constant versus Mass-Preserving Transformation Model

As we showed in the Theorems 6.3.2 and 6.3.4 the two transformation models we presented behave differently: While we can show polyconvexity of the resulting distance measure and therefore existence for the mass-preserving transformation, the resulting distance measure is concave with respect to the Jacobian determinant and thus upper semicontinuous for the intensity-constant transformation operator. Resulting from this different behaviour of the transformation operators are the following two problems:

1. Can we show the existence of a minimizer for the intensity-constant transformation operator, despite the resulting distance measure being concave with respect to the Jacobian determinant?

2. Can we observe the different (analytical) behaviour of both transformation operators also in numerical experiments?

We start with presenting some ideas on how to overcome the difficulties in the analysis of the resulting registration functional. Since the volume term in the hyperelastic regularization is convex, it is a possibility to increase the weight of this term until this convexity dominates the concavity of the Jacobian determinant. This would allow us to show existence of a minimizer for a certain range of regularization parameters but on the other hand eventually discard the possibility to choose the regularization parameter in dependence of the noise level of the data.

The next possibility is to show existence in the sense of (orientation-preserving) Young measures, which we mentioned shortly in Section 3.4. According to [86, Thm. 6.1] every infimizing sequence v_j has a subsequence v_{j_k}, such that the gradients of this subsequence ∇v_{j_k} generate a Young measure, that minimizes the registration functional in the sense of Young measures. It is not directly clear, whether this measure is a Dirac-measure.

The next possibility is to evaluate, whether existence results can be stated, although the functional is not (poly)convex. This question has inter alia been covered by Pietro Celada et al. e.g. in [31, 30]. The central idea for a functional J of the type

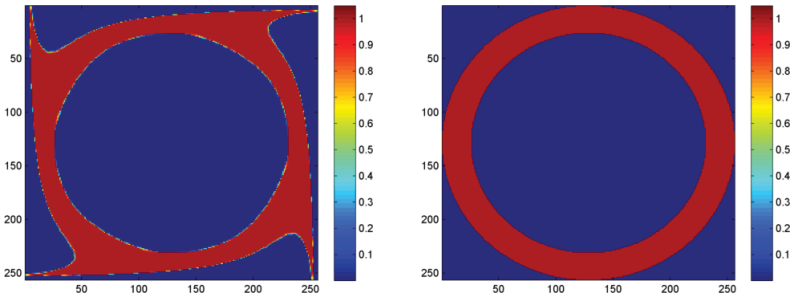
$$J(y) = \int_\Omega g(y(x), \nabla y(x), x) \mathrm{d}x$$

is to study the following functional, generated by the convex envelope g^{**}:

$$J^{**}(y) := \int_\Omega g^{**}(y(x), \nabla y(x), x) \mathrm{d}x.$$

Since J^{**} is polyconvex, existence of a minimizer for J^{**} is guaranteed. If the convex envelope is affine on each connected component of the so called detachment set $\{g^{**} < g\}$ and fulfils some regularity conditions (continuity and growth conditions), then existence of a minimizer is granted [30, Thm. 2.1]. Note that in order to use this result we have to compute the convex envelope for our registration functional, which is a difficult task. However, since the distance measure is concave with respect to the Jacobian determinant, the convex envelope for the distance measure is given by the constant zero function. This might indicate, that the lack of convexity is too grave to obtain an existence result following Celada's reasoning.

In order to study the numerical behaviour we study the following registration problem:

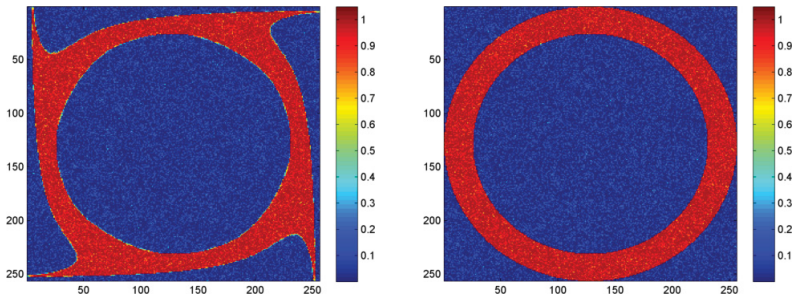


(a) Reference (b) Template

Figure 6.29: Artificial registration problem - the total amount of intensity is the same in each picture.

Since both images hold the exact amount of intensity and take only values in $[0, 1]$ we expect that they can be aligned exactly with the mass-preserving transformation operator as well as the intensity constant. In order to compare both approaches we

corrupt these images with Gaussian noise with mean zero and standard deviation $\sigma = 0.01$:



(a) Reference (b) Template

Figure 6.30: Noisy version of the artificial registration problem in Figure 6.34. Images corrupted by Gaussian noise with $\mu = 0$ and $\sigma = 0.01$.

We start by registering these images for a wide parameter range with the mass-preserving Gaussian as well as the intensity constant Gaussian distance. This leads to the following phantom matching errors:

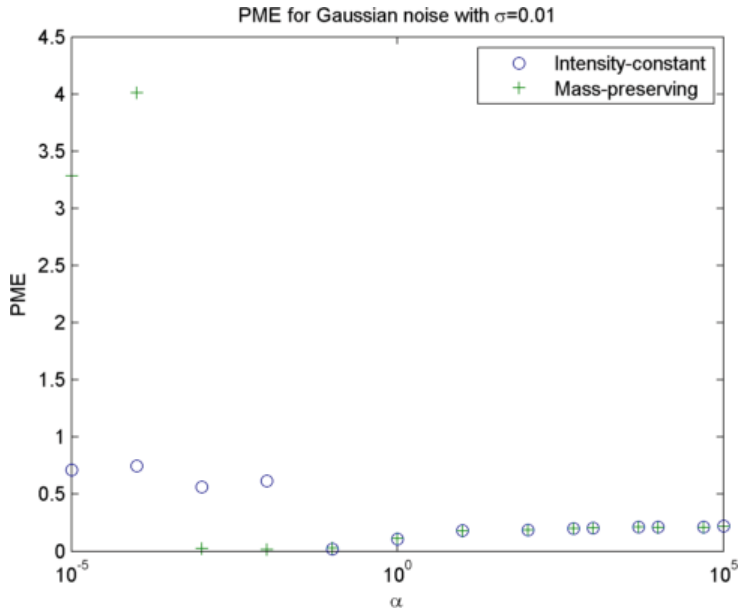


Figure 6.31: Phantom matching errors for the mass-preserving and intensity-constant Gaussian distance. The mass-preserving transformation transformation operator leads to smaller phantom matching error for two values.

We see that there is a small parameter range, where the mass-preserving transformation operator performs better than the intensity-constant in terms of the phantom matching error. Interestingly this small parameter range is also visible in the regularization energies for the estimated transformations.

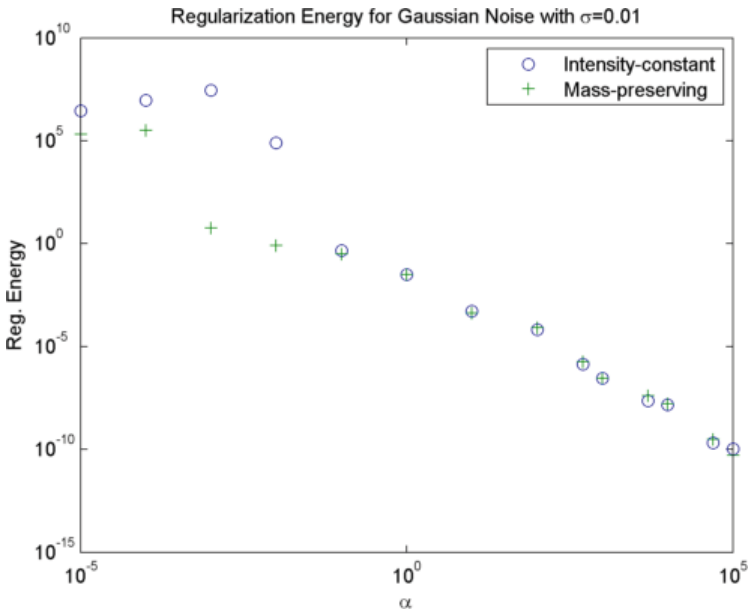
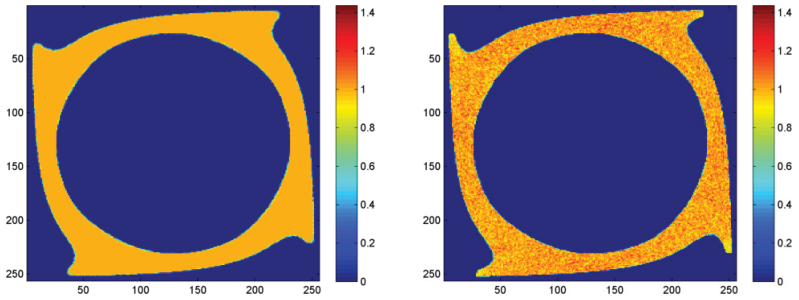


Figure 6.32: Regularization energy of the estimated transformations for mass-preserving and intensity-constant transformation error. The regularization energy is significantly lower in the corresponding parameter range to Figure 6.35

Having compared both approaches in a quantitative way, we turn our focus towards visual inspection.



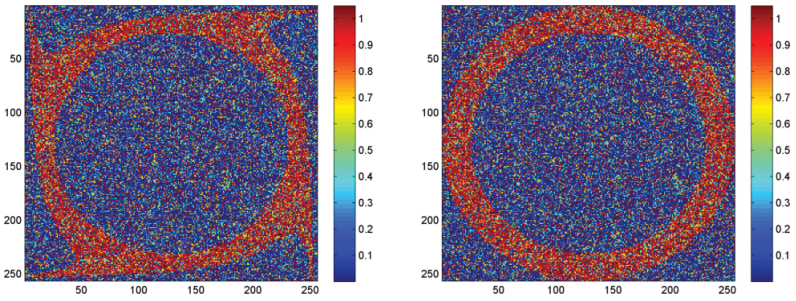
(a) Intensity-constant $\alpha = 0.1$

(b) Mass-preserving $\alpha = 0.01$

Figure 6.33: Template transformed with the transformation best in the sense of the phantom matching error for both methods. The mass-preserving transformation error captures the wings in the corners of the image better than the intensity-constant transformation operator.

We see that the pointed wing shapes on the corners of the image are captured slightly better by the mass-preserving transformation operator. Note that because of this, the phantom matching error is slightly smaller, despite the modulation by multiplication with the Jacobian determinant causing errors in the intensity value.

To conclude this examination, we investigate, whether this properties can also be observed for a higher noise corruption. Thus we corrupt the images in Figure 6.34 with heavier noise.



(a) Reference (b) Template

Figure 6.34: Noisy version of the artificial registration problem in Figure 6.34. Images corrupted by Gaussian noise with $\mu = 0$ and $\sigma = 0.025$.

Again we inspect the phantom matching error for both transformation models.

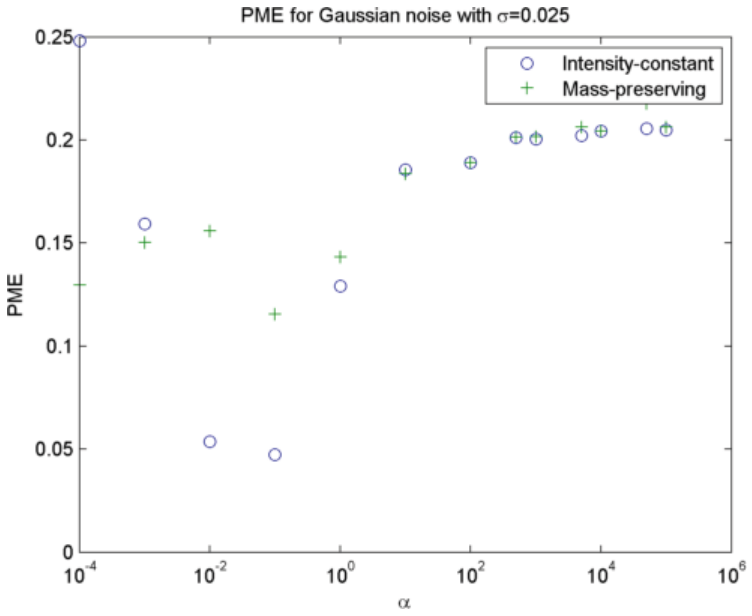


Figure 6.35: Phantom matching errors for the mass-preserving and intensity-constant Gaussian distance. For the high noise corruption the intensity-constant transformation operator leads to transformations with significantly lower phantom matching errors.

Interestingly for the higher noise corruption there is a parameter range, where the intensity-constant transformation modeling performs better than the mass-preserving in terms of the phantom matching error. Again we turn our focus towards the regularization energy of the estimated transformations.

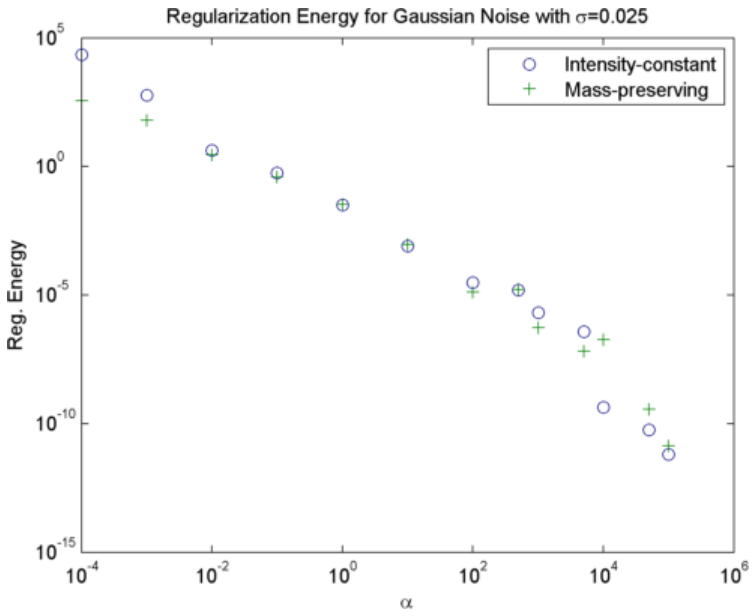
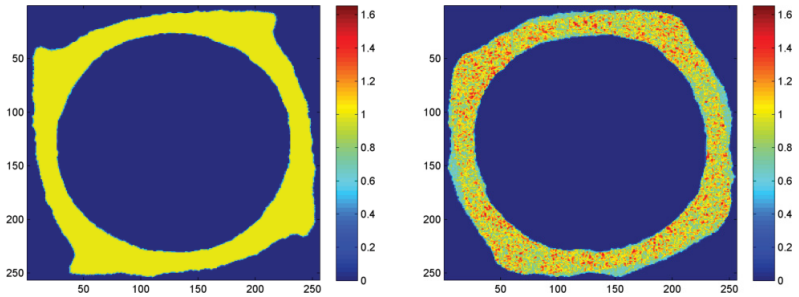


Figure 6.36: Regularization energy of the estimated transformations for mass-preserving and intensity-constant transformation error. There is no parameter range where the regularization energies differ significantly for both transformation models.

In contrast to the case of a low noise corruption (compare Figure 6.32) we cannot observe a significant difference in the regularization energies for the transformation models corresponding to the parameters, where the intensity-constant operator performs better. To conclude we inspect ground truth template transformed by the optimal transformation visually:



(a) Intensity-constant $\alpha = 0.1$

(b) Mass-preserving $\alpha = 0.1$

Figure 6.37: Template transformed with the transformation best in the sense of the phantom matching error for both methods. While the intensity-constant transformation operator rudimentary captures the pointed wings in the corners of the image, the mass-preserving transformation operator fails to do so.

As we observe for a higher noise corruption the intensity-constant transformation operator captures the wings in the corners of the image at least rudimentary, while the mass-preserving transformation operator just slightly deforms the ring. An explanation could be that the pointed wings are interpreted as noise artifacts, which get smoothed by the mass-preserving transformation operator.

To conclude we can say that the mass-preserving transformation operator performed better than the intensity-constant for a low noise corruption. The better performance was clearly visible in the plots of phantom matching error (Figure 6.35) and the regularization energy (Figure 6.32). Increasing the noise gave different results; the intensity-constant transformation operator now performed better than the the mass-preserving. While this could be explained with implicit smoothing properties of the mass-preserving operator, further research on this matter needs to be carried out.

7 Motion-Corrected Reconstruction

This chapter deals with the reconstruction of density images from indirect measurements, such as occuring for example in emission tomography. Since the emission tomography projection operator, representing the scanner geometry, is often compact and the measurements are noise corrupted, direct inversion is not an option (compare Example 4.2.1). For the specific case of Poisson noise, Shepp and Vardi proposed the famous and widely used EM-Algorithm [137], which was extended to regularized reconstruction (see for example [22]). Motion correction can be performed via gating of the images and performing registration on the reconstructed gates. By averaging the registered images the signal-to-noise ratio of one gate could be improved [63]. Recently combined approaches for motion-correction and reconstruction received more and more attention: While the first step in this process was to incorporate independently estimated motion information into the reconstruction process [117], the latest approaches tend to a simultaneous estimation of the reconstructed density image and motion vector field [96, 18]. In the following we propose a framework for motion corrected reconstruction with appropriate regularization for the image (TV) as well as the motion vectors (hyperelastic).

7.1 Problem Formulation

We are concerned with the reconstruction of a sequence of density images $\rho^0, ..., \rho^N$, where each ρ^i corresponds to a data gate f^i. Since we reconstruct the distribution of a radioactive tracer, we expect mass-preservation:

$$\int_{\Omega_0} \rho^0(x)\mathrm{d}x = \int_{\Omega_i} \rho^i(x)\mathrm{d}x \qquad \forall 1 \leq i \leq N \tag{7.1}$$

This guarantees that neither radio activity is created nor vanishing between two gates. Thus we assume the ρ^i to be of the following form:

$$\rho^i(x) := \rho^0(y^i(x)) \det(\nabla y^i)(x), \tag{7.2}$$

where $\rho^0 : \Omega_0 \to \mathbb{R}$ is the so called reference configuration and we set $y^0 := Id$. Furthermore we assume each transformation $y_i : \Omega_i \to \Omega_0$ to be defined on a specific domain Ω_i. Note that we allow these domains to be different, but in many practical applications we have in fact $\Omega_0 = \Omega_i$ for all $1 \leq i \leq N$. By the area formula (Theorem 3.2.7) we obtain mass-preservation for weak diffeomorphisms. The gated data f^i is a noisy version of the projected density $K\rho^i$ with a compact operator $K : \Omega \to \Sigma$, projecting from the image domain Ω into the measurement domain Σ with integration measure $\mathrm{d}\sigma$. In order to

solve this motion-corrected reconstruction problem, we extend the Bayesian framework for reconstruction we presented earlier in this thesis. Consequently we aim to maximize the probability

$$P((\rho, y)|f). \tag{7.3}$$

Bayes' theorem now gives us

$$P((\rho, y)|f) \propto P(f|\rho)P(\rho|y)P(y). \tag{7.4}$$

We will enlighten this three probabilities now further. We start by setting

$$P(f|\rho) = \prod_{i=0}^{N} L(K\rho^i, f^i). \tag{7.5}$$

Here L is the likelihood, adapted to the noise model we assume for the measurement process. Since the focus of our application lies on photon emission tomography, we assume Poisson noise [125]. Nevertheless the analysis presented later on can be extended to different noise characteristics. For the probability of the image sequence with given deformations we choose

$$P(\rho|y) \propto P_0(\rho^0) \prod_{i=1}^{N} \varepsilon(\rho^i - \rho^0(y^i) \det(\nabla y^i)) P_i(\rho^i). \tag{7.6}$$

Here ε is the concentrated measure, which enforces (7.2). Additionally we choose $P(y)$ and $P(\rho)$ as suitable priors for motion vectors and density images. Since we aim to find a maximum a-posteriori probability estimator (MAP-estimator), we minimize the negative log-likelihood of (7.4):

$$\min_{\rho,y} \sum_{i=0}^{N} (D(K\rho^i, f^i) - \log(P_i(\rho^i))) - \log(P(y)), \tag{7.7}$$

where $D(K\rho^i, f^i)$ is the negative log-likelihood adapted to the noise characteristics. We choose the negative logarithms of the priors related to regularization functionals presented earlier. Since we aim to reconstruct images with sharp edges, we use a total variation prior:

$$-\log(P_i(\rho^i)) = \alpha_k |\rho^i|_{BV(\Omega)} \tag{7.8}$$

For the transformation we need a regularization controlling the Jacobian determinant $\det(\nabla y^i)$, in order to enforce the mass-preservation (7.1) with our transformation model (7.2). Thus we choose the hyperelastic regularization energy presented earlier

$$-log(P(y)) = \sum_{i=1}^{N} \beta_i S^{hyper}(y^i). \tag{7.9}$$

Note that the hyperelastic regularization may still lead to non-injective transformations (compare Example 3.2.2), which would violate the mass-preservation condition (7.1) for

the given transformation model (7.2). However we will ignore this fact for the moment and discuss it later.
With the given priors our motion-corrected reconstruction problem is then given by

$$\begin{aligned}\min_{\rho^0,y} J(\rho,y) = &\sum_{i=0}^{N}(D(K(\rho^0(y^i)\det(\nabla y^i)), f^i) + \alpha_i|\rho^i|_{BV(\Omega)}) + \sum_{i=1}^{N}\beta_i S^{hyper}(y^i)\\ &\text{s.t. } \rho^0 \geq 0.\end{aligned} \tag{7.10}$$

Since our main application is the reconstruction of cardiac PET, we specifically present the data fidelity for this case:

$$D(K\rho^i, f^i) = \int_\Sigma K\rho^i - f^i\log(K\rho^i)\mathrm{d}\sigma \tag{7.11}$$

Here $\mathrm{d}\sigma$ denotes the Lebesgue measure for the measurement domain. Having presented a detailed modeling of the motion-corrected reconstruction problem, we turn our focus to the analysis of the resulting variational problem.

7.2 Analysis

The goal of this section is the presentation of an analytical framework for the motion-corrected reconstruction problem. We will show the existence of a minimizer via the direct method in the calculus of variations: We have to verifiy two properties of our functional:

(lsc) Lower semicontinuity

(csls) Compactness of sublevel sets resp. coercivity

While the coercivity can be obtained from the regularizer and lower bounds of the distance measure, the proof of lower semicontinuity needs some caretaking.
As a first step in the analysis of the problem we start by defining the function space, in which we aim to find a minimizer of our functional. The TV-regularization determines that we are looking for a density image in $BV(\Omega)$, while the hyperelastic regularization provides transformations in the subset of admissible transformations $\mathscr{A} \subset H^1$. Consequently we aim to find a minimizer $(\hat{\rho}, \hat{y})$ in the space

$$\mathscr{X} := BV(\Omega) \times (\mathscr{A})^N. \tag{7.12}$$

In order to use the results by Ruthotto we also use the splitting he introduced [129] with the following auxiliary space

$$\tilde{\mathscr{X}} := BV(\Omega) \times H^1(\Omega;\mathbb{R}^d) \times L^4(\Omega;\mathbb{R}^{d\times d}) \times L^2(\Omega), \tag{7.13}$$

equipped with the norm

$$\|(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y))\|_{\tilde{\mathcal{X}}} := \|\rho\|_{BV(\Omega)} + \|y\|_{H^1} + \|\operatorname{cof}(\nabla y)\|_{L^4(\Omega;\mathbb{R}^{d\times d})} + \|\det(\nabla y)\|_2, \tag{7.14}$$

where the norms of the transformations are understood componentwise as

$$\|y\|_{H^1} := \sum_{i=1}^{N} \|y^i\|_{H^1} \tag{7.15}$$

and for cofactor and determinant analogously. In favor of keeping the following proofs as simple as possible, we avoid to use the splitting of the transformations as long as possible and use this argumentation only when it is inevitable. Having stated the preliminaries on the function spaces and norms we turn our focus on the projection operator: Since our proof for the lower semicontinuity follows Resmerita and Anderssen [122], we impose similar assumptions on the projection operator K:

(A1) The operator $K : L^1(\Omega) \to L^1(\Sigma)$ is linear and compact.

(A2) The operator K satisfies $Kx > 0$ a.e. for any $x > 0$ a.e..

(A3) Given $g \in \Sigma, g \geq c_2 > 0$, then $K^*g \geq c_3 > 0$ a.e..

As the analysis of the Kullback-Leibler functional by Resmerita and Anderssen [122] is only done for stationary reconstruction, we need to do some work before we can use the results presented in their paper. Thus we present our crucial result about the sequence of transformed images, which will allow us to show lower semicontinuity following a reasoning with help of Fatou's Lemma [60, Chapter 1, Theorem 2 (iii)]:

Theorem 7.2.1. *Let $\rho_k^0 \rightharpoonup^* \rho^0$ in $BV(\Omega_0)$, $y_k \rightharpoonup y$ in $H^1(\Omega)$ and $\det(\nabla y_k) \rightharpoonup \det(\nabla y)$ in $L^2(\Omega)$. Assume additionally, that*

$$\sup_k \|N_{y_k}(\cdot, \Omega)\|_\infty \leq C \in \mathbb{R}. \tag{7.16}$$

Then we obtain:

$$\rho_k^0(y_k) \det(\nabla y_k) \rightharpoonup \rho^0(y) \det(\nabla y) \text{ in } L^1(\Omega).$$

Proof. For the ease of presentation we use the following abbreviation:

$$\rho_k := \rho_k^0 \quad \rho := \rho^0 \quad d_k := \det(\nabla y_k)$$

For any fixed $\varphi \in (L^1)^* = L^\infty$ we show:

$$0 = \lim_{k\to\infty} \int_\Omega (\rho_k(y_k)d_k - \rho(y)d)\varphi \mathrm{d}x$$

$$= \lim_{k\to\infty} \int_\Omega (\rho_k(y_k)d_k - \rho(y_k)d_k + \rho(y_k)d_k - \rho(y)d)\varphi \mathrm{d}x$$
$$= \lim_{k\to\infty} \int_\Omega (\rho_k(y_k)d_k - \rho(y_k)d_k)\varphi \mathrm{d}x + \lim_{k\to\infty} \int_\Omega (\rho(y_k)d_k - \rho(y)d)\varphi \mathrm{d}x$$

We examine both terms separately and show each of them converges to 0.
We recall that weak star convergence in BV implies strong convergence in L^1 (Definition 4.2.4). Now the first term is straightforward by the area formula and with using (7.16)

$$\lim_{k\to\infty} \int_\Omega |\rho_k(y_k)d_k - \rho(y_k)d_k| \, \mathrm{d}x$$
$$= \lim_{k\to\infty} \int_{\mathbb{R}^n} |\rho_k - \rho| \, N_{y_k}(x, \Omega) \mathrm{d}x$$
$$\leq \lim_{k\to\infty} C \int_{\mathbb{R}^n} |\rho_k - \rho| \, \mathrm{d}x$$

Since $\mathrm{supp}(\rho_k - \rho) \subseteq \Omega_0$ and $\rho_k \to \rho$ in $L^1(\Omega_0)$ implies weak convergence, we obtain

$$\lim_{k\to\infty} \int_\Omega (\rho_k(y_k)d_k - \rho(y_k)d_k)\varphi \mathrm{d}x = 0 \quad \forall \varphi. \tag{7.17}$$

The second term needs more caretaking: According to Theorem 4.2.5 we find a sequence of functions $(\xi^n)_n \subset C^\infty(\Omega_0)$ with

$$\lim_{n\to\infty} \|\xi^n - \rho\|_1 = 0.$$

Let $\epsilon > 0$. Thus we can pick a fix N, such that:

$$\int_{\Omega_0} |\xi^N - \rho| \mathrm{d}x \leq \frac{\epsilon}{4\|\varphi\|_\infty C}. \tag{7.18}$$

We now expand the first term with said ξ^N and obtain:

$$\int_\Omega (\rho(y_k)d_k - \xi^N(y_k)d_k + \xi^N(y_k)d_k - \xi^N(y)d_k$$
$$+ \quad \xi^N(y)d_k - \xi^N(y)d + \xi^N(y)d - \rho(y)d)\varphi \mathrm{d}x$$
$$= \int_\Omega (\rho(y_k)d_k - \xi^N(y_k)d_k)\varphi \mathrm{d}x + \int_\Omega (\xi^N(y_k)d_k - \xi^N(y)d_k)\varphi \mathrm{d}x$$
$$+ \int_\Omega (\xi^N(y)d_k - \xi^N(y)d)\varphi \mathrm{d}x + \int_\Omega (\xi^N(y)d - \rho(y)d)\varphi \mathrm{d}x,$$

We examine each summand separately:
Summand 1:
We start with applying Hölder's inequality and obtain:

$$\begin{aligned}
&\int_\Omega (\rho(y_k)d_k - \xi^N(y_k)d_k)\varphi \mathrm{d}x \\
\leq &\int_\Omega |(\rho(y_k)d_k - \xi^N(y_k)d_k)\varphi| \mathrm{d}x \\
\leq &\|\rho(y_k)d_k - \xi^N(y_k)d_k\|_1 \|\varphi\|_\infty \\
= &\int_\Omega |\rho(y_k)d_k - \xi^N(y_k)d_k| \mathrm{d}x \|\varphi\|_\infty \\
= &\int_{\mathbb{R}^n} |\rho - \xi^N| N_{y_k}(z,\Omega) \mathrm{d}z \|\varphi\|_\infty \\
\leq &C\|\rho - \xi^N\|_1 \|\varphi\|_\infty.
\end{aligned}$$

Combining this with (7.18) we obtain:

$$\int_\Omega (\rho(y_k)d_k - \xi^N(y_k)d_k)\varphi \mathrm{d}x \leq \frac{\epsilon}{4}. \tag{7.19}$$

Summand 2:
First we need a basic observation about this integral:

$$\int_\Omega |\xi^N(y_k)d_k| \mathrm{d}x \leq \|\xi^N(y_k)\|_2 \|d_k\|_2. \tag{7.20}$$

By using that $\xi^N(y_k)$ is bounded, Ω compact and y_k is weakly convergent, it directly follows, that $\|\xi^N(y_k)\|_2 < \infty$ and $\|d_k\|_2$ is bounded by some $E \in \mathbb{R}$ (Banach-Steinhaus, see e.g. [4, Chapter 5, Theorem 3]). Thus (7.20) is finite. Aswell we can directly deduce

$$\int_\Omega |\xi^N(y)d_k| \mathrm{d}x \leq \|\xi^N(y)\|_2 \|d_k\|_2 \leq \infty. \tag{7.21}$$

We can compute now:

$$\begin{aligned}
&\int_\Omega (\xi^N(y_k)d_k - \xi^N(y)d_k)\varphi \mathrm{d}x \\
\leq &\|(\xi^N(y_k) - \xi^N(y))\varphi\|_2 \|d_k\|_2
\end{aligned}$$

$$\leq \|(\xi^N(y_k) - \xi^N(y))\varphi\|_2 E$$
$$= \sqrt{\int_\Omega ((\xi^N(y_k) - \xi^N(y))\varphi)^2 \mathrm{d}x} E$$

Note that ξ^N is C^∞ and therefore $\left(\xi^N\right)^2$ is Lipschitz-continuous with some constant L, since Ω is compact. We obtain

$$\int_\Omega \left((\xi^N(y_k) - \xi^N(y))\varphi\right)^2 \mathrm{d}x \leq \left(\|\varphi\|_\infty\right)^2 L\|y_k - y\|_2. \tag{7.22}$$

By the Rellich-Kondrachov compactness theorem [45, Chapter 5.7, Theorem 1] y_k converges strongly to y in L^2 and so we find $K_2 \in \mathbb{N}$, such that

$$\|y_k - y\| < \frac{\epsilon^2}{16\left(\|\varphi\|_\infty\right)^2 LE^2}. \tag{7.23}$$

This implies for each $k \geq K_2$

$$\int_\Omega (\xi^N(y_k)d - \xi^N(y)d)\varphi \mathrm{d}x < \frac{\epsilon}{4}. \tag{7.24}$$

Summand 3:
Since $\xi^N(y_k)$ is bounded it follows from the weak convergence of the determinants, that there exists $K_3 \in \mathbb{N}$, such that for every $k \geq K_3$

$$\int_\Omega \left(\xi^N(y)d_k - \xi^N(y)d\right)\varphi \mathrm{d}x \leq \frac{\epsilon}{4}. \tag{7.25}$$

Note that the compactness of Ω ensures that $\varphi\xi^N(y) \in L^2$.

Summand 4:
With K_1 like in the explation to summand 1, the same argument yields:

$$\int_\Omega \left(\xi^N(y)d - \rho(y)d\right)\varphi \mathrm{d}x \leq \frac{\epsilon}{4} \tag{7.26}$$

for any $k \geq K_1$.
By combining (7.19), (7.24), (7.25) and (7.26), we obtain for every

$$k \geq K := \max\{K_1, K_2, K_3\}$$

$$\int_\Omega (\rho(y_k)d_k - \rho(y)d)\varphi \mathrm{d}x \leq \epsilon \quad \forall\varphi$$

and therefore

$$\lim_{k\to\infty} \int_{\Omega} (\rho_k(y_k)d_k - \rho_k(y)d)\varphi \mathrm{d}x = 0. \tag{7.27}$$

The assertion follows by combining (7.17) and (7.27). □

Remarks and Extensions. The boundedness of the Banach indicatrix (7.16) is crucial for the proof - but it is not obvious that this is true. To ensure this bound we could either seek for an injective solution (which is numerically challenging) or appoint Dirichlet boundary conditions ($y_k|_{\partial\Omega} = Id$), which leads to injective solutions as well, but is easier to handle in a numerical framework.
Another approach could be to establish bounds for the Banach indicatrix, by taking the regularization energy into account.

With that at hand, we can observe lower semicontinuity of our functional:

Lemma 7.2.2 (Lower Semicontinuity for the Kullback-Leibler data fidelity)**.** *Let the previous assumptions from Theorem 7.2.1 be fulfilled. Then our functional J, defined by*

$$D(\rho_0, y) = \sum_{i=0}^{N} \int_{\Sigma} \underbrace{K(\rho^0(y^i)det(\nabla y^i)) - f^i \log(K(\rho^0(y^i)det(\nabla y^i)))}_{:=g^i(K\rho^0(y^i)\det(\nabla y^i))} \mathrm{d}\sigma \tag{7.28}$$

is lower semicontinous with respect to weak-star convergence in $BV(\Omega_0)$, weak convergence in $W^{1,2}(\Omega_i)$ for the transformations and weak convergence in $L^2(\Omega_i)$ for the determinants.

Proof. Let $y_k \rightharpoonup y$ in $W^{1,2}(\Omega)$, $\det(\nabla y_k) \rightharpoonup \det(\nabla y)$ in L^2, $\rho_k^0 \rightharpoonup \rho^0$ in $BV(\Omega)$.
Since we denote $y = (y^1, ...y^N)$ as the collection of all transformations we understand the weak convergence componentwise.
We recall Theorem 7.2.1 and see that for any fixed $1 \leq i \leq N$

$$\rho_k^0(y_k^i)det(\nabla y_k^i) \rightharpoonup \rho^0(y^i)det(\nabla y^i). \tag{7.29}$$

Since K is a compact operator, by Proposition 3.5.13 K is completely continuous, which gives us:

$$K(\rho_k^0(y_k^i)det(\nabla y_k^i)) \to K(\rho^0(y^i)\det(\nabla y^i)) \qquad \text{in } L^1(\Sigma). \tag{7.30}$$

Therefore we can follow the proof of Lemma 3.4 (iii) from [122] for each summand and obtain lower semicontinuity of the Kullback-Leibler data fidelity term with the following reasoning:
Since $K(\rho_k^0(y_k^i)det(\nabla y_k^i)) \to K(\rho^0(y^i)\det(\nabla y^i))$, we have convergence almost everywhere. Thus we can deduce

$$g^i(K\rho_k^0(y_k^i)\det(\nabla y_k^i)) \to g^i(K\rho^0(y^i)\det(\nabla y^i)) \qquad \text{in } L^1(\Sigma). \tag{7.31}$$

Now we can apply Fatou's Lemma [60, Chapter 1, Theorem 2 (iii)] and obtain:

$$\int_\Sigma g^i(K\rho^0(y^i)\det(\nabla y^i))\mathrm{d}\sigma \leq \liminf_{k\to\infty} \int_\Sigma g^i(K\rho_k^0(y_k^i)\det(\nabla y_k^i))\mathrm{d}\sigma \tag{7.32}$$

Having shown lower semicontinuity for an arbitrary summand with fixed i, the assertion follows directly.

□

Remarks and Extensions. The lemma holds not only for the Kullback-Leibler divergence, but for any distance

$$D(K\rho, f) = \sum_{i=0}^{N} \int_\Sigma g^i(K\rho^i, f^i)\mathrm{d}\sigma \tag{7.33}$$

satisfying

$$g^i(K\rho_k^i, f^i) \to g^i(K\rho^i, f^i) \qquad \text{in } L^1(\Sigma), \tag{7.34}$$

whenever

$$K\rho_k^i \to K\rho^i \qquad \text{in } L^1(\Sigma). \tag{7.35}$$

We have now stated lower semicontinuity results for a wide range of distance terms, including the Kullback-Leibler data fidelity. Thus we turn our focus to the TV-regularization. By setting $\alpha_k = 0$ for any $1 \leq k \leq N$ in (7.10), this would follow directly by the properties of the TV-seminorm we mentioned earlier. However to overcome this restriction we give the following lemma:

Lemma 7.2.3. *Let $\rho_n \rightharpoonup \rho$ in $L^1(\Omega)$. Then*

$$|\rho|_{BV(\Omega)} \leq \liminf_{n\to\infty} |\rho_n|_{BV(\Omega)}. \tag{7.36}$$

Proof. The proof follows [26]. Let $g_l \in C_0^\infty(\Omega\mathbb{R}^d)$ with $\|g_l\|_\infty \leq 1$ satisfying

$$|\rho|_{BV(\Omega)} = \lim_{l\to\infty} \int_\Omega \rho\nabla \cdot g_l \mathrm{d}x. \tag{7.37}$$

Since $\nabla \cdot g \in L^\infty(\Omega)$, we can deduce

$$\int_\Omega \rho\nabla \cdot g\mathrm{d}x = \lim_{n\to\infty} \int_\Omega \rho_n\nabla \cdot g\mathrm{d}x. \tag{7.38}$$

This leads to

$$\begin{aligned}\int_\Omega \rho\nabla \cdot g_l\mathrm{d}x &= \lim_{n\to\infty} \int_\Omega \rho_n\nabla \cdot g_l\mathrm{d}x \\ &= \liminf_{n\to\infty} \int_\Omega \rho_n\nabla \cdot g_l\mathrm{d}x\end{aligned}$$

$$\leq \liminf_{n\to\infty} \sup_{\substack{g\in C_0^\infty(\Omega\mathbb{R}^d)\\ \|g\|_\infty\leq 1}} \int_\Omega \rho_n \nabla\cdot g \mathrm{d}x$$
$$= \liminf_{n\to\infty} |\rho_n|_{BV(\Omega)}.$$

Applying the limit for l leads to

$$TV(\rho) \leq \liminf_{n\to\infty} |\rho_n|_{BV(\Omega)}. \tag{7.39}$$

□

We now have verified the first condition for the existence of a minimizer, so we can turn our focus on the coercivity properties.

Lemma 7.2.4 (ρ-Coercivity for the Kullback-Leibler divergence)**.** *Let our assumptions* (4.68) *for our transformation and (A1)-(A3) for the operator be fulfilled. Then*

$$J^1(\rho_0, y) := \sum_i \int_\Sigma K(\rho_0(y_i det(\nabla y))) - f\log(K(\rho_0(y_i det(\nabla y))))\mathrm{d}\sigma + \alpha_i |\rho_0(y_i)\det(\nabla y_i)|_{BV(\Omega)}$$

is coercive with respect to the variable ρ_0.

Proof. We begin by observing:

$$\log(x) \leq \frac{1}{a}x + \log(a) \qquad \forall x > 0$$

for any fixed $a > 0$. We add the constant $f\log(f) - f$, such that each summand is non-negative. Note that $y_0 = Id$ and therefore we can bind the Kullback-Leibler divergence from below by:

$$\begin{aligned}
&\sum_{i=0}^N \int_\Sigma K(\rho_0(y_i)det(\nabla y_i)) - f\log(K(\rho_0(y_i)det(\nabla y_i))) + f\log(f) - f\mathrm{d}\sigma\\
\geq &\int_\Sigma K(\rho_0(y_0)det(\nabla y_0)) - f\log(K(\rho_0(y_0)det(\nabla y_0))) + f\log(f) - f\mathrm{d}\sigma\\
= &\int_\Sigma K\rho_0 - f\log(K\rho_0) + f\log(f) - f\mathrm{d}\sigma\\
\geq &\int_\Sigma K\rho_0 - f\left(\frac{1}{f+1}K\rho_0 + \log(f+1)\right) + f\log(f) - f\mathrm{d}\sigma\\
\geq &\int_\Sigma \left(1 - \frac{f}{f+1}\right) K\rho_0 \mathrm{d}\sigma \underbrace{- \int_\Sigma f(f+1) + f\log(f) - f\mathrm{d}\sigma}_{:=c_2\in\mathbb{R}}
\end{aligned}$$

$$
\begin{aligned}
&= \int_\Omega K^* \left(1 - \frac{f}{f+1}\right) \rho_0 \mathrm{d}x + c_2 \\
&\geq \int_\Omega c_1 \rho_0 \mathrm{d}x + c_2 \mathrm{d}x \\
&= c_1 \|\rho_0\|_1 + c_2
\end{aligned}
$$

Therefore we can conclude

$$
J_1(\rho_0, y) \geq c_1 \|\rho_0\|_1 + |\rho_0|_{BV(\Omega)} \geq \min\{c_1, 1\}(\|\rho_0\|_1 + |\rho_0|_{BV(\Omega)}). \tag{7.40}
$$

□

Remarks and Extensions. The lemma holds not only for the Kullback-Leibler divergence, but for any distance

$$
D(K\rho^i, f^i) = \int_\Sigma g^i(K\rho_k) \mathrm{d}\sigma \tag{7.41}
$$

which is bounded below and satisfies

$$
g^0(K\rho^0) \geq c_1 K\rho^0 + c_2 \tag{7.42}
$$

with constants $c_1 > 0$ and $c_2 \in \mathbb{R}$, only dependent of f.

We have proved that the first part of our functional is coerciv in ρ_0, so it remains to be shown, that the hyperelastic regularization is coercive with respect to y. Fortunately this has already been done in [129, Chapter 2, Lemma 1]. Combining these results, we can finally formulate an existence result:

Theorem 7.2.5 (Existence of a minimizer in motion-corrected reconstruction)**.** *Let our assumptions (A1)-(A3) hold, furthermore we assume that we have*

$$
J(1, (Id)^N) < \infty, \tag{7.43}
$$

aswell as the following assumption on the Banach indicatrix:

$$
\|N_y(\cdot, \Omega)\|_\infty \leq c_1 S^{hyper}(y). \tag{7.44}
$$

Then the functional

$$
J(\rho, y) = \sum_{i=0}^{N} (D(K(\rho^0(y^i) \det(\nabla y^i)), f_k) + \alpha_i |\rho^i|_{BV(\Omega)}) + \sum_{i=1}^{N} \beta_i S^{hyper}(y^i) \tag{7.45}
$$

with a distance term fulfilling (7.34) *and* (7.42) *has at least one minimizer* $(\hat{\rho}, \hat{y}) \in BV(\Omega) \times \mathscr{A}^N$*. Particularly this holds for the Kullback-Leibler divergence as distance term.*

Proof. Our proof follows the direct methods in the calculus of variations: We start by looking for a minimizer in the space

$$\tilde{\mathcal{X}} := BV(\Omega) \times (H^1(\Omega;\mathbb{R}^d) \times L^4(\Omega;\mathbb{R}^{d\times d}) \times L^2(\Omega))^N, \tag{7.46}$$

equipped with the norm

$$\|(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y))\|_{\tilde{\mathcal{X}}} := \|\rho\|_{BV(\Omega)} + \sum_{i=1}^{N} \|y^i\|_{H^1} + \|\operatorname{cof}(\nabla y^i)\|_{L^4(\Omega;\mathbb{R}^{d\times d})} + \|\det(\nabla y^i)\|_2. \tag{7.47}$$

We define the functional now on this space in the canonical way by setting:

$$J(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y)) := J(\rho, y)$$

and gather some factors in the Cartesian product by setting $y := (y^1, ..., y^N)$ and for the cofactors and determinants analogously.
Now we deduce by Lemma 7.2.4 and the coercivity of the hyperelastic regularizer, that

$$J(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y)) \geq C(\|\rho\|_{BV} + \|y\|_{H^1} + \|\operatorname{cof}(\nabla y)\|_{L^4(\Omega;\mathbb{R}^{d\times d})} + \|\det(\nabla y)\|_2) \tag{7.48}$$

with C the minimum of the constants in the coercivity estimations.
We fix now any E, such that $J(1, (Id)^N) \leq E$. By the estimation above (7.48) we have that for all tupels $(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y))$ in the sub level set

$$J_E := \{(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y)) | J(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y)) \leq E\} \tag{7.49}$$

the norm is bounded by

$$\|(\rho, y, \operatorname{cof}(\nabla y), \det(\nabla y))\|_{\tilde{\mathcal{X}}} \leq \frac{E}{C}. \tag{7.50}$$

Consequently, each summand in the norm $\|\cdot\|_{\tilde{\mathcal{X}}}$ of the product space is bounded by $\frac{E}{C}$. Thus we can deduce by the Banach-Alaoglu theorem that each factor of the Cartesian product in the level set is contained in a compact set. Now the level set (7.49) is contained in a product of compact sets, which is again compact.
We pick now an infimizing sequence $(\rho_n, y_n, \operatorname{cof}(\nabla y_n), \det(\nabla y_n))$. Then there exists an index N such that

$$(\rho_n, y_n, \operatorname{cof}(\nabla y_n), \det(\nabla y_n)) \in J_E \qquad \forall n \geq N. \tag{7.51}$$

Since J_E is contained in a compact set, there exists a convergent subsequence of this infimizing sequence which we denote again by

$$(\rho_n, y_n, \operatorname{cof}(\nabla y_n), \det(\nabla y_n)) \in J_E \qquad \forall n \geq N.$$

According to Ruthotto ([130] resp. Theorem 4.3.4) this convergence is weakly for his splitting, such that

$$y_n^i \rightharpoonup y \qquad \text{in } H^1(\Omega)$$

$$\begin{aligned}\operatorname{cof}(y_n^i) &\rightharpoonup H = \operatorname{cof}(\nabla y^i) \quad \text{in } L^4(\Omega;\mathbb{R}^{d\times d})\\ \det(y_n^i) &\rightharpoonup h = \det(\nabla y^i) \quad \text{in } L^2(\Omega)\end{aligned}$$

for each $1 \leq i \leq N$.
The convergence in the in $BV(\Omega)$ is deduced by the Banach-Alaoglu theorem and therefore in the weak-star topology:

$$\rho_n^0 \rightharpoonup^* \rho^0 \quad \text{in } BV(\Omega).$$

Now the boundedness of the Banach indicatrix is granted by (7.44) and we can derive for each summand in the first sum by Theorem 7.2.1 and Lemma 7.2.2 and 7.2.3

$$\begin{aligned}D(K(\rho^0(y^i)\det(\nabla y^i)), f_k) + \alpha_i|\rho^i|_{BV(\Omega)}) \leq\\ \liminf_{n\to\infty} D(K(\rho_n^0(y^i)\det(\nabla y_n^i)), f^i) + \alpha_i|\rho^i|_{BV(\Omega)}).\end{aligned} \tag{7.52}$$

Furthermore Ruthotto [129] showed the lower semicontinuity of the hyperelastic regularization. Since a sum of lower semicontinuous functionals is again lower semicontinuous we can deduce the existence of a minimizer in $\tilde{\mathcal{X}}$, and J has a minimizer in $BV(\Omega)\times\mathscr{A}^N$. □

We made two additional assumptions in order to prove the theorem: While the assumption (7.43) should not be critical in practical applications, it is not clear, whether (7.44) holds. Since the area formula gives us for the transformed density image:

$$\int_{\Omega_i} \rho^i \mathrm{d}x = \int_{\Omega_i} \rho^0(y^i(x))\det(\nabla y^i)(x)\mathrm{d}(x) = \int_{\mathbb{R}^d} \rho^0(z) N_{y^i}(z,\Omega)\mathrm{d}z \tag{7.53}$$

we can only expect mass-preservation for transformations, which are injective almost everywhere. Thus we turn our focus on this matter and formulate an similar result for the restriction to injective transformations:

Corollary 7.2.6 (Existence for injective transformations)**.** *Let our assumptions (A1)-(A3) hold, furthermore we assume that we have*

$$J(1,(Id)^N) < \infty. \tag{7.54}$$

Then the functional

$$\begin{aligned}\min_{\rho,y} J(\rho,y) = \sum_{i=0}^{N}(D(K(\rho^0(y^i)\det(\nabla y^i)), f_k) + \alpha_i|\rho^i|_{BV(\Omega)}) + \sum_{i=1}^{N}\beta_i S^{hyper}(y^i)\\ s.\ t.\ N_{y^i}(\cdot,\Omega) \leq 1 \ a.\ e. \quad \text{for } 1 \leq i \leq N\end{aligned} \tag{7.55}$$

with a distance term fulfilling (7.34) *and* (7.42) *has at least one minimizer* $(\hat{\rho},\hat{y}) \in BV(\Omega)\times\mathscr{A}^N$. *Particularly this holds for the Kullback-Leibler divergence as distance term.*

Proof. We can follow the proof of Theorem 7.2.5, but additionally the closedness of the constraint remains to be shown. This is uncritical, since the hyperelastic regularization guarantees us that $\operatorname{cof}(\nabla y_n^i) \in L^4(\Omega;\mathbb{R}^{d\times d})$ for all $n \in \mathbb{N}, 1 \leq i \leq N$. The closedness of this constraint is then directly granted by Theorem 3.3.8. □

While Corollary 7.2.6 provides us a satisfying theoretical result, the injectivity side constraint is difficult to handle in a numerical framework. Since Dirichlet boundary conditions are easier to handle we provide an existence result for a closed subset of the injective transformations:

Corollary 7.2.7 (Existence for Dirichlet boundary conditions)**.** *Let our assumptions (A1)-(A3) hold, furthermore we assume that we have*

$$J(1,(Id)^N) < \infty. \tag{7.56}$$

Then the functional

$$\begin{aligned}\min_{\rho,y} J(\rho,y) = \sum_{i=0}^{N}(D(K(\rho^0(y^i)\det(\nabla y^i)),f_k) + \alpha_i|\rho^i|_{BV(\Omega)}) + \sum_{i=1}^{N}\beta_i S^{hyper}(y^i)\\ s.\ t.\ y^i = v^i \text{ on } \quad \partial\Omega \quad \text{for } 1 \leq i \leq N\end{aligned} \tag{7.57}$$

with injective boundary values v^i in the sense of Sobolev traces and a distance term fulfilling (7.34) *and* (7.42) *has at least one minimizer $(\hat{\rho},\hat{y}) \in BV(\Omega)\times\mathscr{A}^N$. Particularly this holds for the Kullback-Leibler divergence as distance term.*

Proof. Since the boundary conditions are understood in the sense of Sobolev traces, the continuity of the trace operator ensures the closedness of constraint (7.57). Furthermore Proposition 3.2.9 ensures the injectivity of each admissible transformation fulfilling (7.57). Now we can follow the proof of Corollary 7.2.6, resp. Theorem 7.2.5. □

Having stated existence results for the Variational problem of motion-corrected reconstruction we show that the minimization of our functional can be understood as a (nonlinear) regularization method in the sense of Definition 4.1.2:

Theorem 7.2.8. *Let $(f_k^i)_k$ be a sequence of noisy data with*

$$\lim_k f_k^i = f_*^i, \tag{7.58}$$

where f^* is the exact data for an image ρ_* and transformations y_*, such that:

$$(\rho_*,y_*) = \min_{\rho,y}\sum_{i=0}^{N} D(K(\rho(y^i)\det(\nabla y^i)),f_*^i) \tag{7.59}$$

for a non-negative distance D, which is lower semicontinuous in both arguments and fulfilling (7.42)*. Furthermore, we define a sequence of functionals J_k by:*

$$J_k = \sum_{i=0}^{N} (D(K(\rho^0(y^i)\det(\nabla y^i)), f_k^i) + \alpha_k^i |\rho^i|_{BV(\Omega)}) + \sum_{i=1}^{N} \beta_k^i S^{hyper}(y^i) \tag{7.60}$$

Then for $\alpha_k \to 0$ *and* $\beta_k \to 0$ *with*

$$\frac{\sum_{i=0}^{N} D(f_*^i, f_k^i)}{\min_i\{a_k^i, \beta_k^i\}} \to 0 \qquad \frac{\max_i\{a_k^i, \beta_k^i\}}{\min_i\{a_k^i, \beta_k^i\}} \leq C \in \mathbb{R} \forall k \tag{7.61}$$

the sequence $(\hat{\rho}_k, \hat{y}_k)_k$*, with* $(\hat{\rho}_k, \hat{y}_k)$ *being minimizers of* J_k *has a convergent subsequence and the limit* $(\hat{\rho}, \hat{y})$ *fulfills:*

$$\sum_i D(\hat{\rho}(\hat{y}^i)\det(\nabla \hat{y}^i), f_*^i) = \sum_i D(\rho_*(y_*^i)\det(\nabla y_*^i), f_*^i) \tag{7.62}$$

Proof. We assume for the sake of simplicity:

$$K\rho_*^i = f_*^i \tag{7.63}$$

We have for a minimizer $(\hat{\rho}_k, \hat{y}_k)$ of J_k:

$$J_k(\hat{\rho}_k; \hat{y}_k) \leq J_k(\rho_*, y_*) \tag{7.64}$$

Now we can deduce:

$$\sum_{i=0} |\hat{\rho}_k^i|_{BV(\Omega)} + S^{hyper}(\hat{y}_k^i) \leq \frac{\sum_{i=0}^{N} D(K\rho_*^i, f_k^i) + \max_i\{\alpha_k^i, \beta_k^i\}\left(|\rho_*^i|_{BV(\Omega)} + S^{hyper}(y_*^i)\right)}{\min_i\{\alpha_k^i, \beta_k^i\}} \tag{7.65}$$

Using Assumption (7.63) we obtain:

$$\sum_{i=0} |\hat{\rho}_k^i|_{BV(\Omega)} + S^{hyper}(\hat{y}_k^i) \leq \frac{\sum_{i=0}^{N} D(f_*, f_k^i) + \max_i\{\alpha_k^i, \beta_k^i\}\left(|\rho_*^i|_{BV(\Omega)} + S^{hyper}(y_*^i)\right)}{\min_i\{\alpha_k^i, \beta_k^i\}}.$$

Now by the parameter choice (7.61) the right hand side of equation (7.65) is bounded. Since $(\hat{\rho}_k, \hat{y}_k)$ are minimizers of the functional J_k this equation together with the coercivity property of the distance measure grants us boundedness of the norm of $(\hat{\rho}_k, \hat{y}_k)$. Then we can deduce by the Banach-Alaoglu theorem that the sequence $(\hat{\rho}_k, \hat{y}_k)_k$ contains a convergent subsequence, which we denote again by $(\hat{\rho}_k, \hat{y}_k)_k$. For the limit of this sequence $(\hat{\rho}, \hat{y})$ we deduce by the lower semicontinuity of the distance:

$$\sum_{i=0}^{N} D(K\hat{\rho}^i, f_*^i) \leq \liminf_k \sum_{i=0}^{N} D(K\hat{\rho}_k^i, f_k^i)$$
$$\leq \lim_k \sum_{i=0}^{N} D(K\rho_*^i, f_k^i) + \sum_{i=0}^{N} \alpha_k^i |\rho_*^i|_{BV(\Omega)} + \sum_{i=1}^{N} \beta_k^i S^{hyper}(y_*^i)$$

Now the regularization parameters tend to zero and we can deduce:

$$\lim_k \sum_{i=0}^{N} D(K\hat{\rho}^i, f_k^i) + \sum_{i=0}^{N} \alpha_k^i |\rho_*^i|_{BV(\Omega)} + \sum_{i=1}^{N} \beta_k^i S^{hyper}(y_*^i) \leq \sum_{i=0}^{N} D(K\rho_*^i, f_*^i) \tag{7.66}$$

Since $\sum_{i=0}^{N} D(K\rho_*^i, f_*^i) \leq \sum_{i=0}^{N} D(K\hat{\rho}^i, f_*^i)$ by construction the first assertion follows. □

Note that the estimation only holds for the composition and not for the components. Since different transformations can lead to the same transformed image (compare Example 4.3.1) we can in general not expect to derive convergence for both components.

Remarks and Extensions. In order to deal with the non uniqueness of the solution yielded by Theorem 7.2.8 we assume additionally

1. $\lim_k \frac{\sum_{i=0}^{N} D(K\rho_*^i, f_k^i)}{\min_i \{\alpha_k^i, \beta_k^i\}} = 0$,

2. For all α_k^i, β_k^i there exist $\tilde{\alpha}^i = \lim_k \frac{\alpha^i}{\min_i \{\alpha_k^i, \beta_k^i\}}$, resp. $\tilde{\beta}^i = \lim_k \frac{\beta^i}{\min_i \{\alpha_k^i, \beta_k^i\}}$,

then we can deduce

$$\sum_{i=0}^{N} \tilde{\alpha}^i |\hat{\rho}^i|_{BV(\Omega)} + \sum_{i=1} \tilde{\beta}^i S^{hyper}(\hat{y}^i)$$
$$\leq \liminf_k \sum_{i=0}^{N} \tilde{\alpha}^i |\hat{\rho}_k^i|_{BV(\Omega)} + \sum_{i=1}^{N} \tilde{\beta}^i S^{hyper}(\hat{y}_k^i)$$
$$\leq \lim_k \left(\frac{\sum_{i=0}^{N} D(K\rho_*^i, f_k^i)}{\min_i \{\alpha_k^i, \beta_k^i\}} + \sum_{i=0}^{N} \tilde{\alpha}^i |\rho_*^i|_{BV(\Omega)} + \sum_{i=1}^{N} \tilde{\beta}^i S^{hyper}(y_*^i) \right)$$
$$= \sum_{i=0}^{N} \tilde{\alpha}^i |\rho_*^i|_{BV(\Omega)} + \sum_{i=1}^{N} \tilde{\beta}^i S^{hyper}(y_*^i).$$

By using this deduction we can show that $(\hat{\rho}, \hat{y})$ is a solution, which minimizes

$$\sum_{i=0}^{N} \tilde{\alpha}^i |\cdot|_{BV(\Omega)} + \sum_{i=1}^{N} \tilde{\beta}^i S^{hyper}(\cdot) \quad \forall \quad (\rho, y) \quad \text{with} \quad \sum_{i=0}^{N} D(K\rho^i, f_*^i) = 0.$$

This solution can be viewed analogously to the best-approximate solution in the sense of Engl et al. [44, Definition 2.1].

Having stated these analytical properties of our functional, we turn our focus towards the numerical implementation in the next section.

7.3 Numerical Framework

In this section we describe the numerical framework we use to solve the motion-corrected reconstruction problem. Note that we restrict the presentation of the numerical framework to the Kullback-Leibler divergence as distance measure; nevertheless the extension to other distances with given TV-regularized reconstruction algorithms is straightforward.

We aim to perform a First-Discretize-then-Optimize approach combined with an alternating minimization strategy. For this we only need a discretization in the space-domain, since we assume to have time-discretized data. For the time-discretization we assume, that the discrete data $(f_i)_i$ is gated: Therefore we define time nodes $(t_i)_i$ such that each node t_i represents a stage of motion (f.e. cardiac or respiratoric gate). Additionally we impose regularization only on the reference configuration ρ_0 and not on the transformed versions of ρ_0, so by setting $\alpha^i = 0$ for $i > 0$ we obtain the following functional to be minimized

$$\begin{aligned} J(\rho_0, y) = &\sum_i \int_\Sigma K(\rho^0(y^i) det(\nabla y^i)) - f^i \log(K(\rho^0(y^i(det(\nabla y^i))))) \mathrm{d}\sigma \\ &+ |\rho^0|_{BV(\Omega)} + \sum_i S^{hyper}(y^i). \end{aligned}$$

The space discretization is straightforward: We define pairwise disjoint pixel B_i, such that

$$\Omega = \bigcup_i B_i. \tag{7.67}$$

With this discretization at hand we can put our problem in a discrete framework and start to minimize our functional. Given an initial value ρ_0^0 our algorithm reads as follows:

$$\begin{cases} 1.\ \text{Motion step:} & y_{k+1} \in \arg\min_y \{J(\rho_k^0, y)\} \\ 2.\ \text{Reconstruction step:} & \rho_{k+1}^0 \in \arg\min_{\rho^0} \{J(\rho^0, y_{k+1})\} \end{cases}.$$

While the reconstruction step can be realized via motion corrected EM-TV algorithms (see e.g. [22]), the motion step needs some caretaking. In the next sections we want to outline both implementations briefly.

7.3.1 Reconstruction-Step: Motion-Corrected EM-TV

In this section we want to show that we can apply standard EM-TV algorithms to motion-corrected reconstructions as well. For the ease of presentation we assume that our transformations y^i are global invertible. The central idea for EM-TV is to alternate an EM step with given discrete projection operator (so called system matrix) with a TV denoising step [22]. In order to so we will derivate a system matrix for motion-corrected reconstruction in the following, such that we can apply the EM-TV algorithm from [22]. The central idea of the system matrix is to determine the possibility that a photon resulting from a decay in a pixel is detected on a specific line. Thus an element $a_{k,j}$ of the system matrix is the probability that a photon resulting from a decay in pixel j is detected on the line k. It follows:

$$(A\rho)_k = En(L_k),$$

where $En(L_k)$ denotes the expected value of decays detected on the k-th line. Ignoring other effects like scatter we can compute this probability without motion by

$$a_{k,j} = \frac{\int\limits_{L_k} 1_{B_j} \mathrm{d}x}{\int\limits_{L_k} 1 \mathrm{d}x}.$$

In the case of injective motion we aim to obtain a system matrix, which maps the reference acitivity distribution ρ_0 for each time node t_i to the detector domain. Assume that the system matrix A^0 for the reference configuration is known. Inspired by Chapter 6 we can write our transformation model (7.2) equivalently as

$$\rho^i = T^{mp}_{y^i} \rho_0. \tag{7.68}$$

Projecting the transformed reference configuration ρ^i into the measurement domain Σ is then given by

$$A^0 \rho^i = A^0 \left(T^{mp}_{y^i} \rho^0 \right). \tag{7.69}$$

Now using the associative property leads to

$$A^0 \rho^i = \underbrace{\left(A^0 T^{mp}_{y^i} \right)}_{=:A^i} \rho^0, \tag{7.70}$$

which gives us a motion-corrected projection operator A^i, acting on ρ^0 only. Now we can formulate the motion-corrected reconstruction problem as

$$\underbrace{\begin{pmatrix} A^0 \\ A^1 \\ \dots \\ A^N \end{pmatrix}}_{:=A} \rho_0 = \underbrace{\begin{pmatrix} f_0 \\ f_1 \\ \dots \\ f_N \end{pmatrix}}_{:=f}. \tag{7.71}$$

We observe that the structure of (7.71) is like the one in [22] and we can directly apply the presented EM-TV algorithm. Note that one will not store the matrix A explicitly, but focus rather on a parallelized, matrix free implementation.

7.3.2 Motion-Step: Interpretation as Registration

In the Motion-Step we need to minimize

$$\sum_i \int_\Sigma K(\rho_0(y^i)\det(\nabla y^i)) - f_i \log(K(\rho_0(y^i)\det(\nabla y^i)))\mathrm{d}\sigma$$
$$+ |\rho_0|_{BV(\Omega)} + \sum_i S^{hyper}(y^i)$$

with respect to our set of transformations $(y^i)_i$. We notice, that all the summands are independent, so we obtain a minimum of the sum, by minimizing each summand. Since $|\rho_0|_{BV(\Omega)}$ is a constant we retain a problem of the form:

$$\min_{y^i} \int_\Sigma K(\rho_0(y^i)det(\nabla y)) - f_i \log(K(\rho_0(y^i)\det(\nabla y)))\mathrm{d}\sigma + S^{hyper}(y(\cdot,t_i)) \tag{7.72}$$

If we consider the Kullback-Leibler divergence being an distance measure, this is the form of a standrad registration problem from [101] with hyperelastic regularization like we defined in Section 4.3. Nevertheless we should mention that the Kullback-Leibler distance measure is defined on the detector domain Σ and not on the image domain Ω like standard distance measures such as SSD or the Normalized Gradient Field.
We turn our focus towards a minimization framework for (7.72). Since Σ is the detector domain and therefore discrete with size m_Σ, the integral becomes a sum:

$$\sum_{j=1}^{m_\Sigma} K(\rho_0(y^i)\det(\nabla y^i))(j) - f_i(j)\log(K(\rho_0(y^i)det(\nabla y^i))(j)) + S^{hyper}(y(\cdot,t_i)) \tag{7.73}$$

Again we aim to perform a minimization with the modified BFGS method like described in Chapter 5. In order to minimize the objective function (7.73) we need to compute the derivatives with respect to the transformation grid. Like in the previous chapter the transformation is given on a nodal grid:

$$yc \in \mathbb{R}^{\tilde{m}} \qquad \tilde{m} = \prod_{i=1}^{d} (m_i + 1)d, \tag{7.74}$$

where $m = (m_1, .., m_d)$ denotes the size of the digital image obtained in the reconstruction step. Again we use the interpolation inter and computation of the Jacobian determinant jac, implemented in the FAIR toolbox. Then for any continuously differentiable distance term

$$D : (\mathbb{R}^{m_\Sigma})^2 \to \mathbb{R} \qquad (K(\text{inter}(\rho^0, yc)\,\text{jac}(yc)), f) \to D(K(\text{inter}(\rho^0, yc)\,\text{jac}(yc)), f) \tag{7.75}$$

the derivative with respect to the transformation grid yc is given by the chain rule as

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\,yc} D(K(\text{inter}(\rho^0, yc)\,\text{jac}(yc)), f) \\ =&\frac{\mathrm{d}}{\mathrm{d}\,w} D(K(\text{inter}(\rho^0, yc)\,\text{jac}(yc)), f) \left(K \left(\frac{\mathrm{d}}{\mathrm{d}\,yc}\,\text{inter} + \frac{\mathrm{d}}{\mathrm{d}\,yc}\,\text{jac} \right) \right). \end{aligned} \tag{7.76}$$

With the help of (7.76) we can deal with different distance terms in the same objective function. An implementation is given in Section 9.2.2 in the Appendix.
The actual registration is then performed with a multilevel approach like presented in the previous chapter. Since the field of view in the scanner is often bigger than the studied object, it is reasonable to impose Dirichlet boundary conditions on the motion with $y|_{\partial\Omega} = Id$, which guarantees the existence of a minimizer (Corollary 7.2.7). This boundary conditions can be realized by taking the identity as starting guess and modifying the search directions to zero at the boundary.
Having described the numerical framework we use for motion-corrected reconstruction we turn our focus towards applications.

7.4 Results

This section is devoted to the presentation of reconstruction results obtained by the variational approach we outlined in the previous sections. We use three different data sets to highlight the potential gained by our approach. We start with a thorough evaluation of an artificial 2D deblurring problem, followed by some basic tests on the XCAT software phantom as well as the hardware phantom "Wilhelm".

7.4.1 Artificial Deblurring Example

Deblurring problems are often occuring in microscopy, where the exact image gets convoluted with an (often unknown) point-spread function [89, 29]. Since the focus in this section lies on the motion-corrected reconstruction, we assume the exact blur operator to be known. We consider a ring shaped object in three different stages of shrinkage:

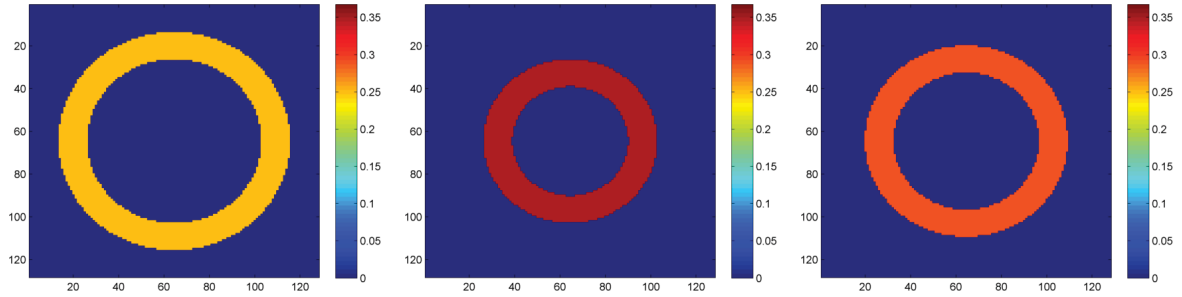


Figure 7.1: Noise free images used to generate data. Note that the amount of intensity stored in each image is the same and they can be aligned exactly with help of an affine linear transformation. We aim to reconstruct the image on the left.

Since we aim to simulate a mass-preserving reconstruction problem, the overall activity stored in each image is the same. In order to keep a simple transformation model, all images can be aligned via 2D affine linear transformations. The noise free data is now generated by applying a 2D disc filter implemented in MATLAB:

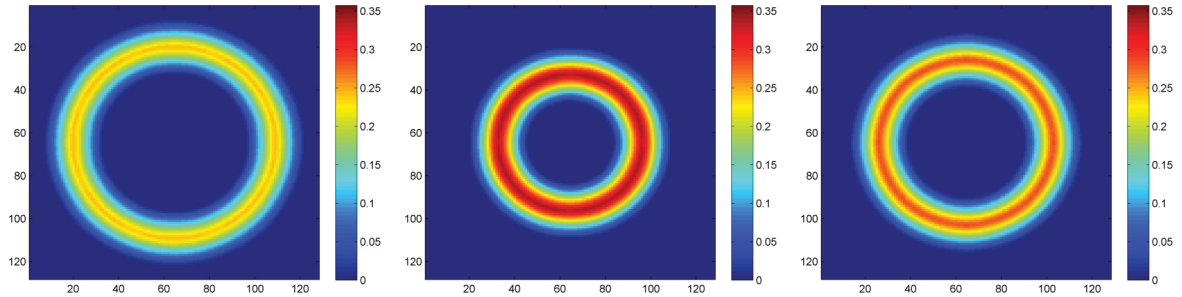


Figure 7.2: Blurred versions of the ground truth images in Figure 7.1. Note that the sharp edges of the images get lost as a consequence of the blurring.

We corrupt these blurred versions now by Poisson noise, in order to simulate the noise corruption occurring in photon counting processes:

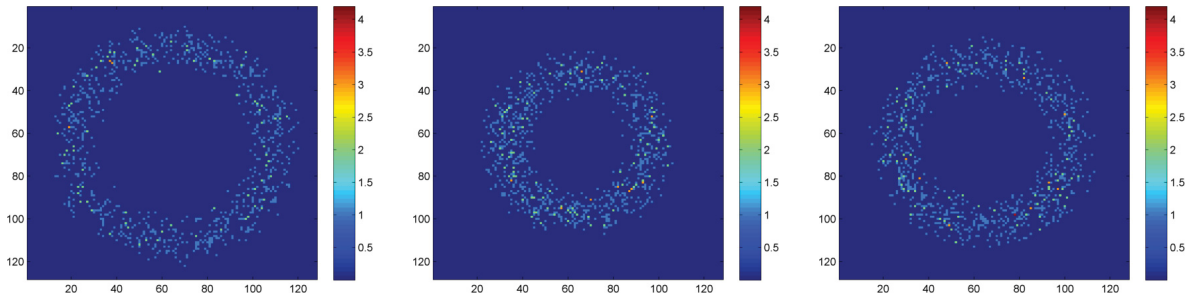


Figure 7.3: Blurred images in Figure 7.2 corrupted by Poisson noise. Note that much of the information is lost.

As we see the ring gets thinned out by the Poisson noise, which causes massive loss of information. We first start by comparing the proposed method with (TV-regularized) single gate reconstructions. In the proposed method we chose an affine linear 2D transformation model with hyperelastic regularization with weights $\beta_1 = \beta_3 = 1$ and $\beta_2 = 0$ and the Kullback-Leibler divergence as data fidelity. In the proposed method as well as the EM-TV reconstruction 5 Bregman iterations (see e.g. [22]) were performed for contrast enhancement. For various TV regularization weights α we obtain the following reconstruction errors:

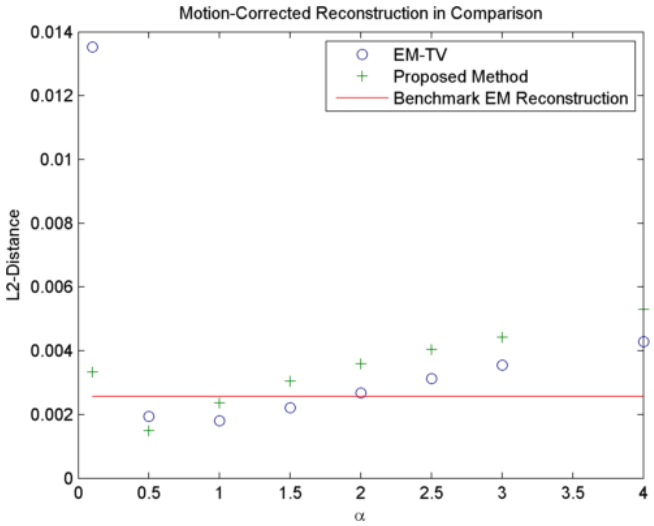
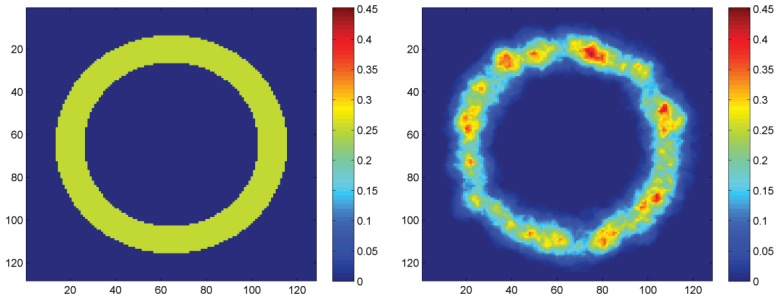


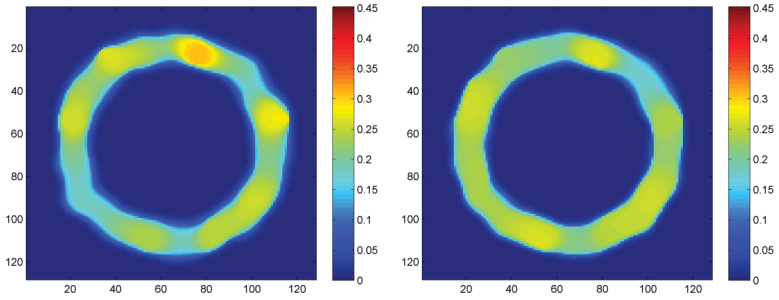
Figure 7.4: Reconstruction error for the proposed Method (5 Bregman Iterations), EM-TV (5 Bregman Iterations) reconstruction and reconstruction error yielded by classical EM as benchmark. The proposed method yields the best reconstruction result.

As we see the proposed method performs slightly better than the TV-regularized expectation maximization, but the reconstruction errors are only slightly different for all three

methods. Thus we turn towards visual inspection of the three reconstruction methods.



(a) Ground truth (b) EM reconstruction



(c) EM-TV reconstruction (d) Proposed method

Figure 7.5: Ground truth and reconstruction results for the three methods. The proposed method nearly recovers the ground truth images, while noise artifacts can be seen in the EM reconstruction and marginally in the TV-regularized EM reconstruction.

By visual inspection we see clearly that the EM reconstruction is noise corrupted. While the contour of the object is recovered, there are several irregularities in the shape as well as the intensity. The TV regularized reconstruction has a much clearer homogeneous structure, but the noise artifacts in the top of the image can still be seen. These artifiacts are nearly completely removed by the proposed method, which results in the nearly uniform recovering of the ring. Thus visually inspection showed us that the proposed method yields the best reconstruction result.

We turn our focus now towards the evaluation of the estimated motion vectors. In order to assess the quality of the estimation we compare the motion estimation by the proposed method with an affine mass-preserving 2D registration performed on TV regularized single gate reconstructions. For doing so, we picked the best reconstruction in the terms of the reconstruction error for the EM-TV reconstruction of each gate and registered them. This leads to the following phantom matching errors:

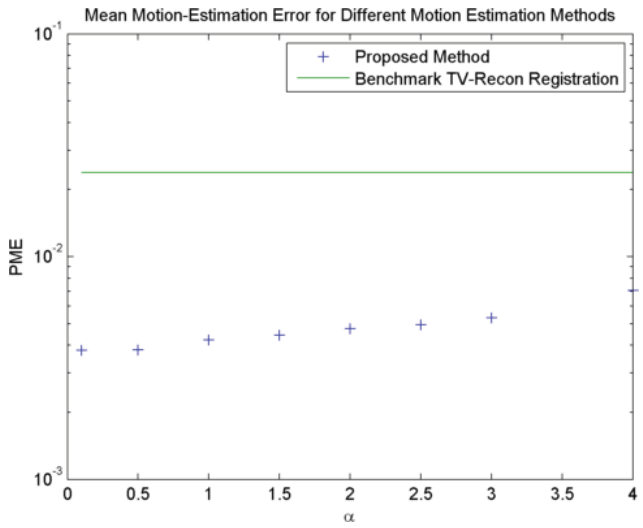


Figure 7.6: Logarithmic plot of the averaged phantom matching errors for transformations yielded by the proposed method and registration performed on the best TV regularized single gate reconstructions as benchmark. The phantom matching errors for all transformations were averaged. The error for the proposed method is smaller by an order of magnitude.

We compare the proposed method now with motion correction techniques using the motion information gained from the affine 2D registration. As we expect incorporating this motion information into the reconstruction process does not produce a better reconstruction than Bayesian ML estimation.

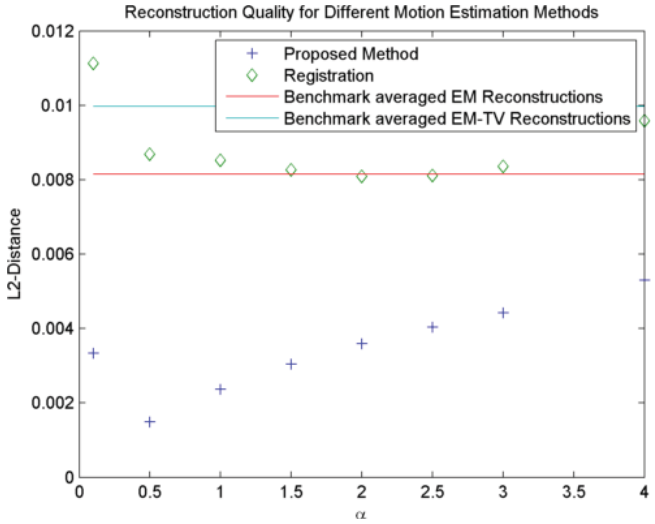
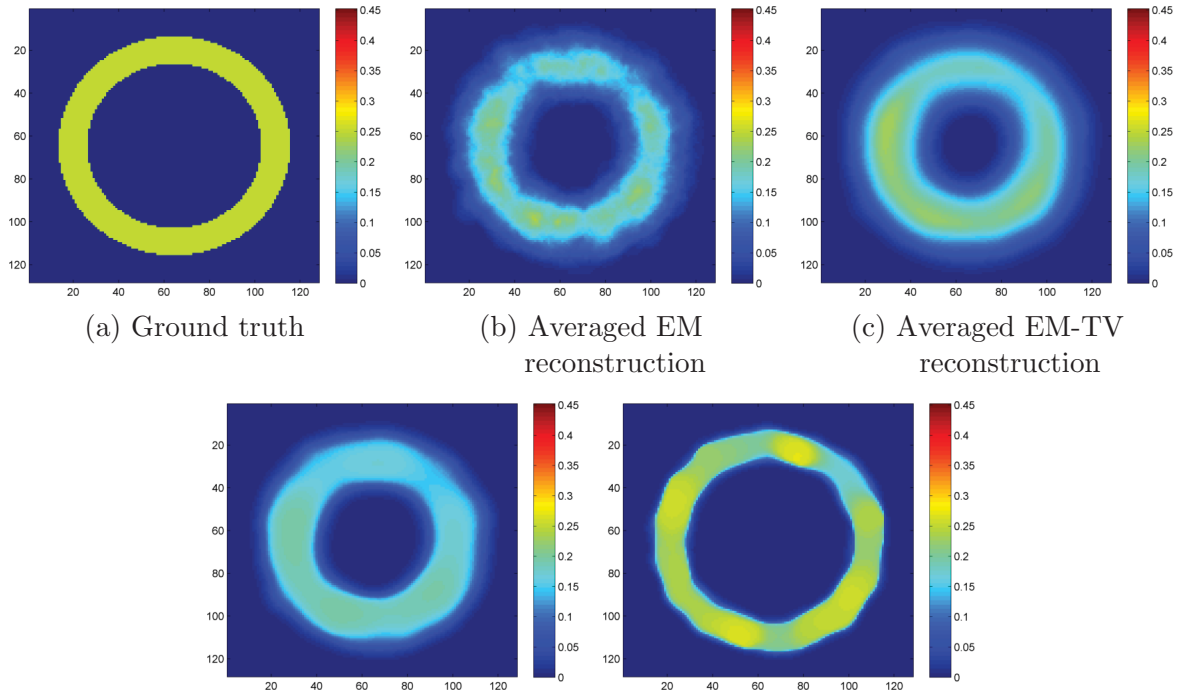


Figure 7.7: Reconstruction error for the proposed method, motion-corrected EM-TV with motion determined by registration of single gate registrations an averaged (TV regularized) single gate reconstructions as benchmark. The other motion correction techniques suffer heavily from the inaccurate motion estimation.

To complete this comparison we visually inspect the reconstruction results in Figure 7.7 for all methods :



(a) Ground truth

(b) Averaged EM reconstruction

(c) Averaged EM-TV reconstruction

(d) Registration based m.c. EM-TV

(e) Proposed method

Figure 7.8: Best (TV-regularized) single gate reconstruction averaged by registration, best motion-corrected EM-TV reconstruction with motion estimation via registration of EM-TV reconstructed single gates and best reconstruction yielded by the proposed methods. Note that the precise size of the ring gets lost due to the inaccurate motion estimation in the registration based methods.

We see that the inaccurate motion estimation by registering TV regularized single gate reconstruction has severe impact on the reconstruction quality. Since the edges are not exactly aligned by the registration, this imprecise motion estimation leads to an inexact size of the ring in the reconstruction, respectively to a blurring of the edges in the averaging process.

In conclusion we can state the advantages of the proposed method are twofold. First incorporating motion-estimation directly is superior to averaging methods, since the reconstruction is performed from full data instead of just averaging images reconstructed from parts of the data (compare Chapter 1). Despite small errors in the motion information the proposed methods also performs better than the single gate reconstruction with TV regularization. Note that while the difference between our method and TV-regularized reconstruction was clearly visible in the visual inspection, both reconstruction methods had a similar L^2-distance to the ground truth image. Additionally the motion-estimation

yielded by the proposed method was clearly superior to the registration based methods. We can give the following three possible explanations for this:

- In contrast to single gate reconstructions the proposed methods transforms a template image, which is generated from the full data set.
- By projecting the transformed template into the measurement domain the error occurring from reconstructing the reference image from part of the data is avoided.
- The Kullback-Leibler data fidelity used in the proposed method is directly adapted to the Poisson noise characteristics of the data, while the SSD distance we use for the registration is used for Gaussian denoising.

Having illustrated the superior behavior of the simultaneous motion and image estimation on this artificial dataset we turn our focus towards real applications.

7.4.2 XCAT Software Phantom Data

In this subsection we inspect the performance of three reconstructions methods on data generated by the XCAT software phantom [134]. For generating the data we projected four cardiac gates of the phantom into the data spaces specified by the Siemens Biograph Sensation 16 scanner provided by the EMRecon toolbox [84]. The projected data was downscaled by the factor 1000 corrupted with Poisson noise and then scaled up again. We reconstructed the first gate from this single gate data with the classical EM algorithm, EM-TV algorithm, and the method we described in this chapter applied on the full data. For various TV-regularization weights α and hyperelastic regularization parameter $\beta_1 = 500\alpha, \beta_2 = 0, \beta_3 = 650\alpha$. Applying these methods with the specified regularization parameters leads to the following reconstruction errors:

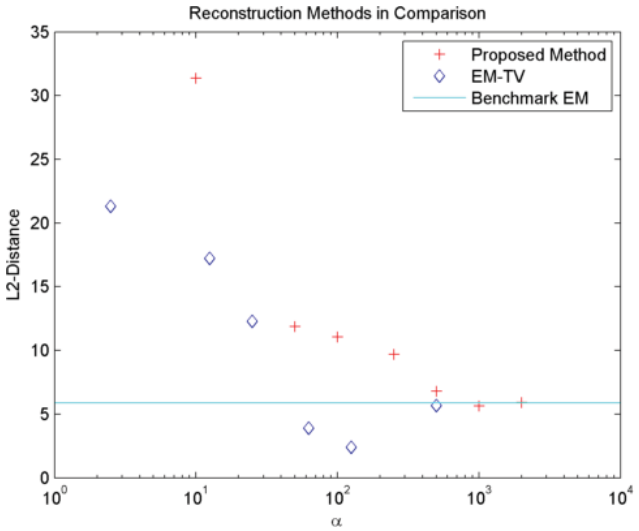
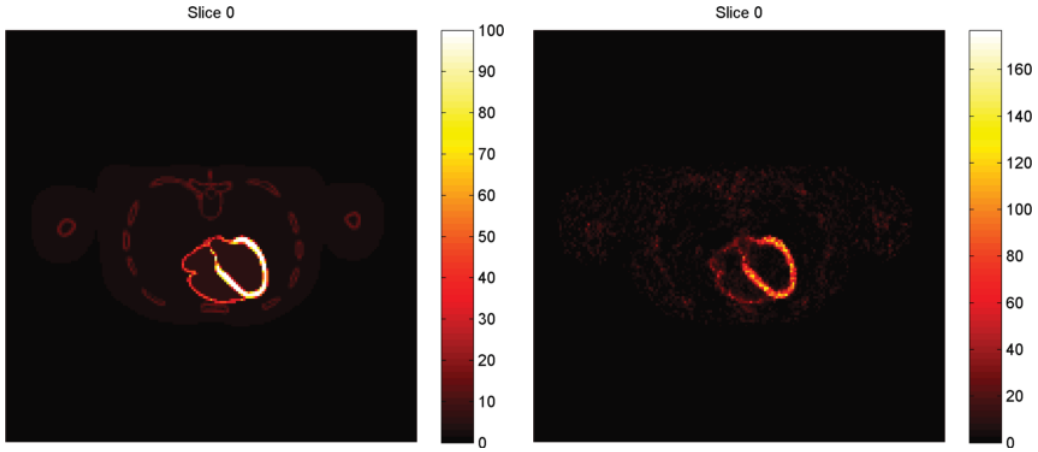
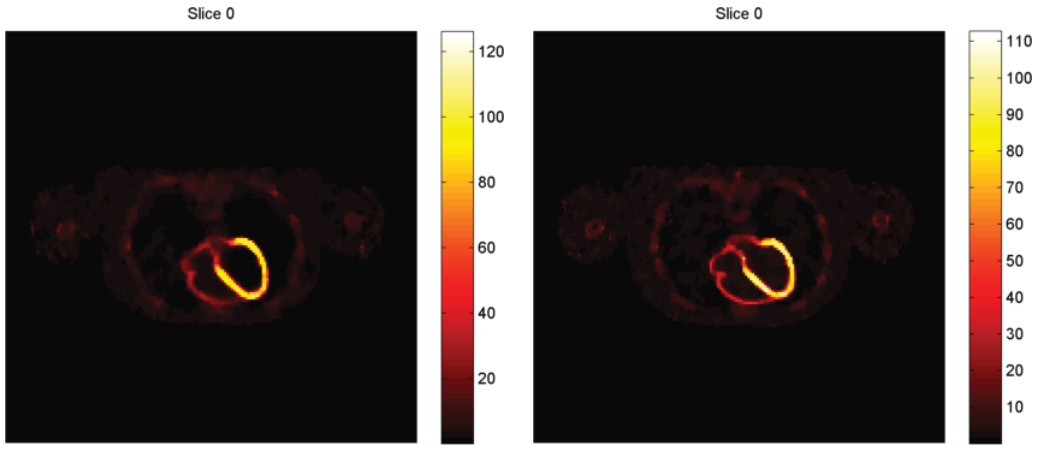


Figure 7.9: Reconstruction errors for EM, EM-TV and the proposed method. EM-TV Reconstruction performs best.

As we see, the EM-TV method seems to perform best in terms of the reconstruction error. Having assessed the quality of the three reconstruction methods quantitatively we turn our focus towards visual inspection of the reconstructed images.



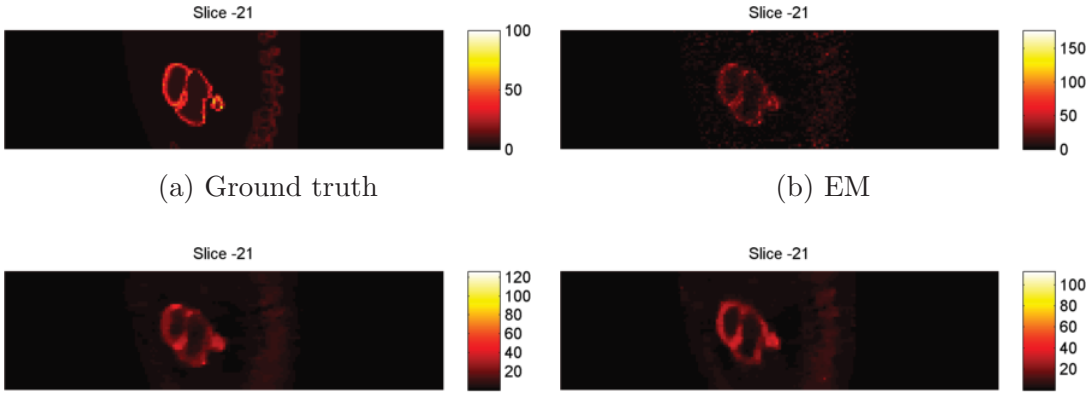
(a) Ground Truth (b) EM



(c) EM-TV (d) Proposed method

Figure 7.10: Ground truth and reconstructions EM, EM-TV ($\alpha = 125$) and the proposed method ($\alpha = 1000$). The TV regularized methods produce clearly better reconstruction results than the classical expectation maximization algorithm. The right ventricle is captured slightly better by the proposed method.

Visual inspection of the reconstructed images shows that despite having a higher reconstruction error the proposed method captures structures with a low amount of activity better than both single gate reconstruction methods. As we see the right ventricle stands out much clearer by incorporating motion information. In order to inspect this phenomenon further, we show some more 2D slices of the reconstructions.



(a) Ground truth (b) EM

(c) EM-TV (d) Proposed method

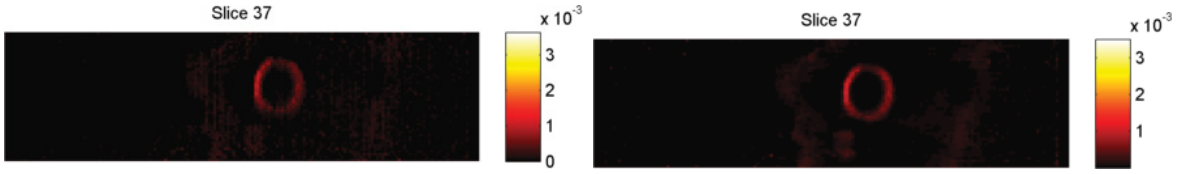
Figure 7.11: Representative 2D slice for the three reconstruction methods. Again the proposed method recovers areas with a low amount of intensity better than the single gate reconstructions.

Despite leading to bigger reconstruction errors than TV-regularized single Gate reconstruction (Figure 7.7) the proposed methods manages to capture areas with low amount of tracer activity better than the single Gate reconstructions (Figures 7.10 and 7.11). This illustrates the potential of motion-corrected reconstruction for images suffering from low signal to noise ratio. As we expected, there are nearly no differences visible in areas with a high amount of tracer activity. The better performance of the TV regularized single gate reconstruction in terms of the reconstruction error can be explained by the fact, that the proposed method suffers from a slightly inaccurate motion-estimation, which leads to the observed reconstruction errors. Thus we state that the proposed method is best suited for images with really poor signal-to-noise ratio, so that the inaccuracies in the motion estimation are dominated by the improved reconstruction using the full data.

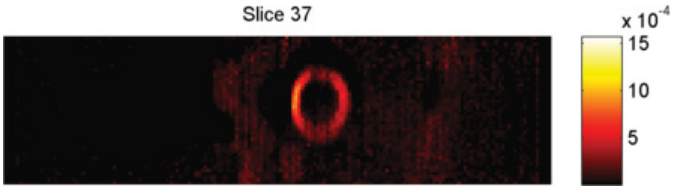
7.4.3 Hardware Phantom "Wilhelm"

To conclude this section we test the proposed method on data generated by the hardware phantom "Wilhelm" [133]. This phantom is built in the torso of a mannequin and consists of a diaphragm, which moves in the imaging process and affects a small flask filled with tracer as well as an artificial left ventricle. By gating we obtained data for eight different stages of motion. We aim to reconstruct one gate with the motion-corrected reconstruction framework (5 Bregman iterations in the reconstruction step) and choose the parameters $\alpha = 200$ for the TV regularizer and $\beta_1 = 50, \beta_2 = 0,$
$\beta_3 = 65$ in the hyperelastic regularization. To compare the results we also reconstruct the single gate with the classical EM algorithm (5 iterations) and a TV-regularized single gate reconstruction ($\alpha = 50$). The TV-regularized reconstructions were performed with 5 EM iterations. Note that we imposed a stronger regularization weight in the motion-corrected approach, since the data term gets bigger due to the comparison with all data sets and not only one.

We start by comparing reconstructions of all three methods in a representative 2D slice picturing the ventricle.



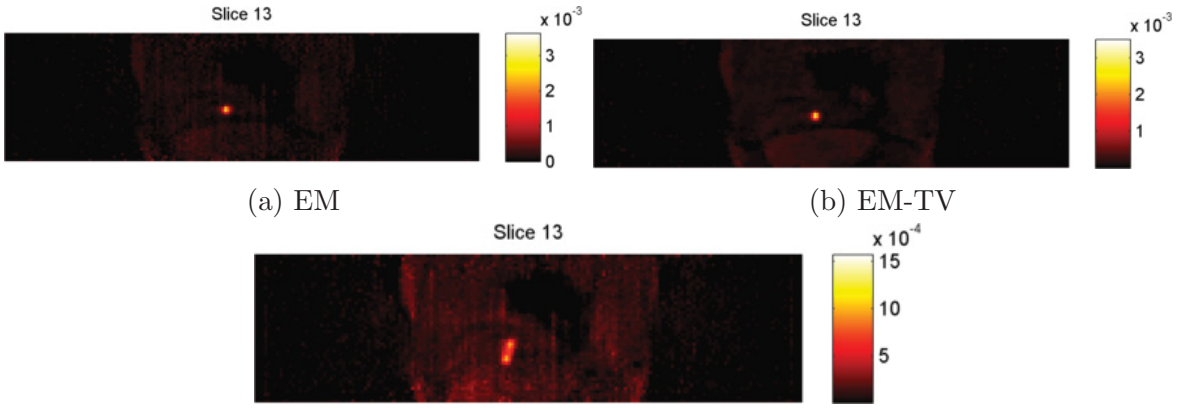
(a) EM (b) EM-TV



(c) Proposed method

Figure 7.12: Representative 2D slice of the reconstructions yielded by the three different reconstruction methods. The proposed methods suffers from loss of contrast, but nevertheless enhances the structure of the ventricle from the background. The reconstruction in the proposed method and the EM-TV algorithm was performed with 5 Bregman iterations.

Note that despite from the loss of contrast the ventricle stands out clearer from the background. As a drawback we observe that the noise in the background gets amplified by the proposed method. A possible explanation for the contrast loss is, that the definition of mass-preservation used in the EMRecon software differs from the notion we used throughout our modeling [85].
We conclude this paragraph by inspecting the reconstruction with focus on the small tracer filled flask moving with the diaphragm.



(a) EM

(b) EM-TV

(c) Proposed method

Figure 7.13: Representative 2D slice of the reconstructions yielded by the three different reconstruction methods. The flask with the tracer is reconstructed nearly without artifacts by the classical methods due to the good signal to noise ratio. In contrast to that the proposed method has difficulties in estimating the large motion of the small flask.

The evaluation of the method on real data showed, that the proposed method yields improvement in the reconstruction of areas with poor signal-to-noise ratio. However there are still two major drawbacks, which need to be addressed in further research. The first is the loss-of-contrast, which we could observe in Figure 7.12, while the second is given by the poor motion estimation for small objects (Figure 7.13).

7.5 Discussion and Outlook

We have presented a novel variational approach to motion corrected reconstruction of density images. After the motivation of the model with help of Bayesian statistics, we proceeded to the analysis of the model. Central part of the analysis was Theorem 7.2.1, which ensured us weak L^1-convergence of the sequence of transformated images, where images as well as transformations were sequences. To prove this theorem we relied heavily on regularity properties granted by the regularizers for intensity and motion vectors. A critical point was the boundedness of the Banach indicatrix in L^∞: It remains unclear, if we can establish a bound of the form

$$\|N_y(\cdot,\Omega)\|_\infty \leq C(\Omega)\mathscr{S}^{hyper}(y). \tag{7.77}$$

In order to give existence results, independent of this assumption, we restricted ourselves to the case of injective transformations. Although this restriction can be motivated by a mass-preservation demand, which is in our model only granted for injective transformations, the numerical realization can again be challenging. We proposed a framework with injective Dirichlet - boundary conditions to guarantee injectivity for the whole domain, but an extension to a less restrictive model might be of interest (see e.g. [6] for a diffeomorphic registration framework).

We thoroughly tested the method on an artificial deblurring example and showed superiority to several other reconstruction methods. By applying the method on soft- and hardware-phantom data we gave a proof-of-concept for the applicability for real data. Although there are still difficulties to deal with the movement of really small objects the method seems to be well-suited for motion-corrected reconstruction of clinical data, especially with a low signal-to-noise ratio. Again a thorough evaluation on clinical data is the next step to go, as well as finding new means to assess the quality of a found transformation in motion correction.
Motion-corrected reconstruction with alternating minimization has been done e.g. by Mair et al. [96], although neither the concept of mass-preservation nor TV regularization is imposed. To the best of our knowledge the work of Blume et al. is most closely related to the presented framework: In [17] the authors propose a similar general framework to reconstruction with simultaneous motion estimation and incorporate a local invertibility constraint in [18]. While the focus of Blume et al. lies on the actual implementation of a reconstruction method with a parametric B-spline transformation model, our main contribution in this chapter is therefore twofold:

- A thorough analysis of the motion-corrected reconstruction problem with mass-preserving transformation model and appropriate regularization for density image and motion.
- Presentation of a numerical framework, capable of solving the designed functional for parametric as well as non-parametric transformation models.

In the following we present some open questions, which can motivate further research in this field.

7.5.1 Weak Formulation of the Reconstruction Problem

In order to present a weaker formulation of the reconstruction problem, we first state some some basic facts about Radon measures before we actually describe the weak formulation.

Radon Measures

We will now state the basic definition and properties for Radon measures from [60, Chapter 1] in order to derive a weaker formulation of the motion-corrected reconstruction problem:

Definition 7.5.1 (Radon Measure)**.** Let μ be a measure over a locally compact and separable metric space $\mathcal{X}$. Furthermore we denote the Borel-Algebra by $\mathscr{B}(\mathcal{X})$ and the sigma-algebra of μ-measurable sets by $\mathscr{M}_\mu$. μ is called a Radon measure, iff

(i) $\mathscr{B}(\mathcal{X}) \subset \mathscr{M}_\mu$.

(ii) For every $A \subset \mathcal{X}$ there exists $B \in \mathscr{B}(\mathcal{X})$ such that $B \supset A$ and $\mu(B) = \mu(A)$.

(iii) For every compact subset $K \subset \mathcal{X}$ we have $\mu(K) < +\infty$.

Remarks and Extensions. A measure μ fulfilling conditions (i) and (ii) is called *Borel-regular*.

Note that the total variation can also be defined for Radon measures (compare [60, Chapter 1.4: Definition 2, Proposition 2]. We will not focus on this matter, but rather state the space of Radon measures can be identified with the dual space of the compactly supported continuous functions:

Theorem 7.5.2. *Let $J : C_c(\mathcal{X}, \mathbb{R})$ be a positive linear functional, i.e. $J(f) \geq 0$, whenever $f \geq 0$. Then there exists a positive Radon measure μ on $\mathcal{X}$, such that*

$$J(f) = \int f \mathrm{d}\mu \qquad \forall f \in C_c(\mathcal{X}). \tag{7.78}$$

Proof. See [60, Chapter 1.4 Theorem 2]. □

Having stated the basic theory, we turn our focus towards presenting a weak formulation for the motion-corrected reconstruction problem. More information on Radon measures can e.g. be found in [60].

Weak Formulation of the Problem

The analysis of functional (7.10) showed, that we needed assumptions on the injectivity resp. boundedness of the Banach indicatrix of the transformation in order to derive analytical results. We want to outline a weak formulation, which guarantees mass preservation, but does not assume injectivity of the transformation. We conclude this outlook by showing that the weak formulation implies injectivity for the motion y and thus corresponds to restricting the admissible set of transformation to a certain subset of the weak diffemorphisms.
We start by giving an equivalent formulation of the transformation model (7.2) for a diffeomorphism y^i with inverse y^i:

$$\int_{\Omega^i} \rho^i(x)\varphi(x)\mathrm{d}x = \int_{\Omega^i} \rho^0(y^i(x)) \det(\nabla y^i(x))\varphi(x)\mathrm{d}x \qquad \forall \varphi \in C_0(\Omega). \tag{7.79}$$

By applying the change of variables formula we obtain

$$\int_{\Omega^i} \rho^i(x)\varphi(x)\mathrm{d}x = \int_{\Omega} \rho^0(x)\varphi(z^i(x))\mathrm{d}x \qquad \forall \varphi \in C_0(\Omega). \tag{7.80}$$

Since $\rho^i \geq 0$ we can relax to Radon measures:

$$\int_{\Omega^i} \varphi(x)\mathrm{d}\mu^i(x) = \int_{\Omega} \rho^0(x)\varphi(z^i(x))\mathrm{d}x \qquad \forall \varphi \in C_0(\Omega). \tag{7.81}$$

Alternatively we can use $\varphi(y^i(x))$ as test function:

$$\int_{\Omega^i} \varphi(y^i(x))\mathrm{d}\mu^i(x) = \int_{\Omega} \rho^0(x)\varphi(x)\mathrm{d}x \qquad \forall \varphi \in C_0(\Omega). \tag{7.82}$$

In order to use (7.81) as a constraint for minimizing (7.10) we provide consistency with the strong formulation by the following propsition:

Proposition 7.5.3. *Let ρ^0 be nonnegative and $z^i \in \mathscr{A}$ an admissible transformation. Then there exists a unique, nonnegative μ^i satisfying* (7.81)*, which fulfils the mass-preservation property*

$$\int_{\Omega^i} \mathrm{d}\mu^i(x) = \int_{\Omega} \rho^0(x)\mathrm{d}x. \tag{7.83}$$

Proof. Note that for any fixed ρ^i and z^i the right hand side of (7.81) defines a linear functional. Now Theorem 7.5.2 ensures the existence of a unique Radon measure μ^i satisfying (7.81). The mass-preservation condition (7.83) can by shown by using test functions converging to constants. □

As a consequence we obtain for the weak formulation of the motion-corrected reconstruction problem:

$$\begin{aligned} \tilde{J}(\mu, z) = &\sum_{i=0}^{N} (D(K\mu^i, f^i) + \alpha^i |\mu^i|_{BV(\Omega)}) + \sum_{i=1}^{N} \beta^i S^{hyper}(z^i) \\ &\text{subject to (7.81)} \end{aligned} \tag{7.84}$$

We will not elaborate further on the analysis for the weak formulation, but instead present a consistency result:

Proposition 7.5.4. *Let $\rho^i \in L^1$ and y^i, z^i fulfil* (7.79) *and* (7.81)*. Then we have $N_{y^i}(\cdot, \Omega) \leq 1$ a.e. and $z^i(x) = (y^i)^{-1}(x)$ for a.e. $x \in \Omega$.*

Proof. We can deduce with the second part of the area formula (Theorem 3.2.7):

$$\begin{aligned} \int_{\Omega^i} \rho^i(x)\varphi(x)\mathrm{d}x &= \int_{\Omega^i} \rho^0(y^i(x)) \det(\nabla y^i(x))\varphi(x)\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \sum_{w \in ((y^i)^{-1}(x) \cap \Omega)} \rho^0(y^i(w))\varphi(w)\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \rho^0(x) \sum_{w \in ((y^i)^{-1}(x) \cap \Omega)} \varphi(w)\mathrm{d}x. \end{aligned}$$

By using Proposition 7.5.3 and (7.81) we obtain:

$$\int_{\mathbb{R}^d} \rho^0(x) \sum_{w\in((y^i)^{-1}(x)\cap\Omega)} \varphi(w)\mathrm{d}x = \int_\Omega \rho^0(x)\varphi(z^i(x))\mathrm{d}x \qquad \forall\varphi\in C_0(\Omega). \tag{7.85}$$

By using test functions φ converging to constants, we can now deduce, that $N_{y^i}(\cdot,\Omega)\leq 1$. It follows that y^i is weakly invertible and we can deduce

$$\int_{\mathbb{R}^d} \rho^0(x)\varphi((y^i)^{-1}(x))\mathrm{d}x = \int_\Omega \rho^0(x)\varphi(z^i(x))\mathrm{d}x \qquad \forall\varphi\in C_0(\Omega). \tag{7.86}$$

Since (7.86) holds for all test functions $\varphi\in C_0(\Omega)$, the assertion follows. □

The proposition above guarantees that a for solution (μ,z) of the weak formulation (7.84), which can be expressed via (7.81) and (7.79), the motion field z^i is the weak inverse of y^i. Whether such an y exists for a given weak solution (μ,z) is object to further research as well as a detailed analysis for the weak formulation.

7.5.2 Attenuation Correction

Attenuation correction plays a big role in the clinical application of emission tomography. See for example [11] for variational approach to obtain motion-corrected attenuation maps. Having ignored this factor in our description we will sketch out shortly how to expand our model to incorporate attenuation information and how this influences analysis and numerics.
In clinical applications one can not expect to have attenuation information on all gates, in fact there is often only one CT image μ available, which corresponds to one gate. In our framework we would choose this gate as the reference gate ρ_0. One often incorporates the information granted by the attenuation map μ into the projection operator [85]. In the stationary case this changes the operator to

$$K(\mu). \tag{7.87}$$

Now we assume that μ is transformed accordingly to the image ρ_0 and we get a attenuation operator incorporating motion information by

$$K(\mu(y)) \qquad \text{or} \qquad K(\mu(y)\det(\nabla y)), \tag{7.88}$$

depending on whether we assume μ being transformed intensity constant or motion corrected. Now the crucial question in the analysis is, if the convergence properties we obtained for a sequence of transformations $y_k\rightharpoonup y$ in $W^{1,2}$ and density images $\rho_k^0\rightharpoonup^*\rho^0$ in $BV(\Omega)$ still hold. Since the CT image μ is a digital image, we can use interpolation techniques to ensure Lipschitz-continuity for μ in the continuous framework (compare [100] for a similar reasoning). The Lipschitz-continuity then ensures strong convergence for the intensity-constant transformation:

$$\mu(y_k)\to\mu(y). \tag{7.89}$$

We use the fact to motivate a convergence of the operator

$$K(\mu(y_k)) \to K(\mu(y)) \tag{7.90}$$

in the norm topology might be achieved. If this convergence is granted, then we can deduce strong convergence of the transformed image $\rho^0(y_k)$ via

$$\begin{aligned} &\|K_k(\rho_k^0(y_k)\det(\nabla y_k)) - K(\rho^0(y)\det(\nabla y))\| \\ \leq &\|(K_k - K)(\rho_l^0(y_k)\det(\nabla y_k))\| + \|K(\rho_k^0(y_k)\det(\nabla y_k) - \rho^0(y)\det(\nabla y))\|, \end{aligned} \tag{7.91}$$

where we used the abbreviations $K := K(\mu(y))$ and $K_k := K(\mu(y_k))$. Now we can obtain a convergence property for $K_k(\rho_k^0(y_k)\det(\nabla y_k))$ with help of estimation (7.91) and the convergence of the operator in the norm topology and the boundedness of $\rho_k^0(y_k)\det(\nabla y_k)$ (first summand in (7.91)) resp. the complete continuity of K and Theorem 7.2.1 (second summand). As this short outline on attenuation correction shows, there are several questions to be adressed in the analysis for attentuation and motion corrected reconstruction:

- Can we verify the convergence $K_k \to K$ in the norm topology for a general class of compact reconstruction operators?
- If we cannot achieve convergence in the norm topology, can we achieve convergence for the operator in weaker topologies (e.g strong operator topology)?
- Can we obtain a strong convergence result for $K_k(\rho_k^0(y_k)\det(\nabla y_k))$ only relying on convergence in weaker topologies?
- How does the weak convergence of the Jacobian determinant influence the convergence properties above, when we assume a mass-preserving transformed attenuation map?

Having identified some central questions for the analysis of the problem we shortly outline the implications in the numerical framework: If we follow the alternating minimization approach we presented, we need to evaluate the consequences for reconstruction and motion step. Since the reconstruction step is performed for a fixed motion, we can use the same implementation of the reconstruction step with an attenuation corrected operator. The motion step gets changed: Since the operator depends now on the motion vectors, we need to compute derivatives for $K(y)(\rho^0(y)\det(\nabla y))$ with respect to y in order to obtain an adapted object function. If this adapted objective function is given, we can follow the minimization approach we presented earlier.

7.5.3 A Priori Information

The framework we presented is only capable of incorporating rather general a priori information via the regularization terms. In practical applications often additional information is available: In the following we discuss, how we can modify our framework to be able to deal with this information and present applications. This outlook is divided into paragraphs for a priori information on image and motion.

Information on the Image

Recently PET-MRI is used in preclinical studies more and more often ([80]). PET-MRI is a hybrid technique performing simultaneous PET and MR imaging, which allows to incorporate a priori information resulting from the MR images into the PET reconstruction. Recently Ehrhardt et al. [41] proposed a method for joint-PET-MRI reconstruction by evaluating the similarity of PET and MRI image via a NGF-type metric. Inspired by this idea, we expand our minimization problem for given MR images u^i to

$$\min_{\rho,y} J(\rho, y) = \sum_{i=0}^{N} (D(K\rho^i, f^i) + D^{NGF}(\rho^i, u^i) + \alpha_i |\rho^i|_{BV(\Omega)}) + \sum_{i=1}^{N} \beta_i S^{hyper}(y^i). \quad (7.92)$$

It is not directly clear, whether our analytical results can be extended to this expanded functional: Although we have the convergence of the composition (Theorem 7.2.1) as a backbone for the analysis, this convergence is only weak in the L^1 sense. Thus the question arising is, if we can show a lower semicontinuity property for the NGF distance relying only on this weak convergence. If we expand this model additionally to a joint-reconstruction, this question becomes even more complex, since we need to study the lower semicontinuity of the NGF measure for sequences in both arguments.

Besides these analytically challenging problems, one has also to deal with a changed numerical framework. Since the summand D^{NGF} influences the reconstruction as well as the motion step both implementations need to be adapted. While the reconstruction step can be performed like described by Ehrhardt [41] with the fixed motion vectors incorporated in the operator (compare Subsection 7.3.1), the motion-step can be handled by solving a registration problem with two distance measures.

Information on the Motion

We will discuss two possible ways how to incorporate a priori motion vectors into the reconstruction process. Motion-corrected reconstruction is sometimes performed with motion information $\tilde{y}$ obtained from registration of MR images [117]. Instead of using this information directly for a reconstruction, we can rather take the registration as an initial guess for the motion step and slightly change the regularizer to penalizing the displacement $y - \tilde{y}$ instead of $y - Id$ in area, surface and volume term.

We motivate the second way with the following example: The moving mice project [97] tries to reconstruct density images for freely moving mice, which are recorded by cameras. Since the mice move in a box inside the PET scanner, the displacements are much bigger than respiratory or cardiac motion for fixated mouses. A possibility to deal with this is to formulate a reconstruction frame work with Dirichlet boundary conditions (compare Corollary 7.2.7) and use the recorded motion of the mouse as a boundary value. This idea can be generalized to the cases where motion on the boundary of a studied object is available as a priori information.

7.5.4 Different Minimization Strategies and Convergence

We did not address convergence properties of the alternating minimization algorithm in the presentation of our numerical framework for the minimization of the functional for motion-corrected reconstruction. While convergence results are available for the (damped) Bregman-EM-TV algorithm [131], we used for the reconstruction steps and can be stated in a similar way as in Section 6.6.2 for the registration problem, solved in the motion step, the convergence of the alternating minimization is an open problem.

Although the algorithm performed well (see Section 7.4), there are few convergence results for applying alternating minimization algorithm to a minimization problem defined on a product space. In [27, Theorem 4.3] Burger and Scherzer showed convergence of the algorithm for a blind deconvolution problem. General results are given with a special focus on adaptive algorithms are given in [108] and the references therein. These results rely on the so called three- and four-point-properties, which are hard to verify for our minimization problem.

To our best knowledge the result most closely related to the minimization problem we face has been given by Beck [15]. In a discrete setting Beck considers the minimization of an objective function of the form

$$H(\rho, y) = D(\rho, y) + g_1(\rho) + g_2(y). \tag{7.93}$$

Beck shows that any accumulation point of the sequence generated by the alternating minimization algorithm is a stationary point, if the following conditions hold:

1. The functions g_1 and g_2 are closed and proper convex functions, assumed to be subdifferentiable.
2. The function D is continuously differentiable.
3. The gradient of D is (uniformly) Lipschitz-continuous with respect to ρ for any fixed y.
4. The gradient of D is Lipschitz-continuous with respect to y for any fixed ρ.
5. The optimal set of (7.93) is nonempty. In addition the problems

$$\min_{\rho} D(\rho, \tilde{y}) + g_1(\rho) \qquad \min_{y} D(\tilde{\rho}, y) + g_2(y)$$

   have minimizers for any fixed $\tilde{y}$, resp. $\tilde{\rho}$.

The first assumption is the most crucial one: While the convexity of the TV-regularization is well known, verifying convexity for the hyperelastic regularization on the motion vectors is a difficult task. A possible subject of further research on this matter is, whether the splitting of the transformation into gradient, cofactor matrix and Jacobian determinant, can again be applied to relax this assumption in order to show the convergence property.

While the assumptions 2 - 5 can be addressed with an interpolation reasoning (compare Section 6.6.2), the last assumption should not be critical in practical applications.
Having outlined an approach to verify a convergence property for the alternating minimization approach we chose in our numerical framework, we conclude this discussion by mentioning that convergence properties can be stated for other minimization strategies: Recently the primal-dual algorithm by Chambolle and Pock [32] has received more and more attention in image processing. Despite being initially presented for convex problems, extensions to non-convex problems have been given (compare e.g [111]). A proximal regularization of the alternating minimization introduced by Attouch et al. ([7]) might even better suited for the discrete objective function (7.93). For a given initial ρ their algorithm is given by:

$$\begin{cases} y_{k+1} \in \arg\min_{y} \left\{ H(\rho_k, y) + \frac{1}{2\lambda_k} \|y - y_{k+1}\|^2 \right\} \\ \rho_{k+1} \in \arg\min_{\rho} \left\{ H(\rho, y_{k+1}) + \frac{1}{2\mu_k} \|\rho - \rho_k\|^2 \right\} \end{cases}$$

According to the authors this algorithm converges to a stationary point, provided the functions g_1 and g_2 are proper and lower semicontinuous and D is continuously differentiable with Lipschitz-continuous gradients. Additionally the so called Kurdyka-Lojasiewicz inequality needs to be satisfied (compare [7, Section 3.2]). Implementing this proximal regularized algorithm as well as verifying the Kurdyka-Lojasiewicz inequality for a wide range of data fidelities is object to further research.

8 Concluding Discussion

This thesis was devoted to the analysis and numerical implementation of two mathematical problems resulting from simultaneous image and motion estimation. Although the problems presented in Chapters 6 and 7 were developed to accomplish different goals they are nevertheless closely related: While an implicit image estimation enhanced the quality of image registrations on noisy image sequences in Chapter 6, we used motion-estimation to obtain better reconstruction results in Chapter 7. Both resulting functionals can be viewed as special cases of the functional

$$F(u,v) = \sum_{i=0}^{N} D(K(T_{v^i}u), f^i) + \alpha^i R_I(T_{v^i}u) + R_M(v^i). \tag{8.1}$$

We obtained the functionals for registration with noise modeling by setting

$$K = Id \qquad \alpha^i = 0 \quad \forall \quad 0 \leq i \leq N. \tag{8.2}$$

Note that we restricted the presentation of this functional to the case $N = 2$, but the extension is straightforward. In contrast the functionals for motion-corrected reconstruction are obtained via setting

$$T_{v^i}u := u(v)\det(\nabla v). \tag{8.3}$$

Consequently the analysis of both functionals relied on the convergence properties of sequences of the type

$$u_n(v_n) \qquad \text{resp.} \qquad u_n(v_n)\det(\nabla v) \tag{8.4}$$

for sequences u_n and v_n converging weakly in certain spaces depending on the regularization. Key tool in studying this convergence properties was the area formula (Theorem 3.2.7), which we presented for the admissible transformations given by the hyperelastic regularization. Since we imposed no additional regularization on the image u in Chapter 6, the composition $u_n(v_n)$ can be viewed as the product of two weakly-convergent sequences, which converges not necessarily (Theorem 6.3.7). The TV-regularization we imposed for motion-corrected reconstruction granted a strong convergence in L^1, which helped us to prove weak convergence of the composition (Theorem 7.2.1). The complete continuity of the projection operator K then ensured strong convergence of the projected composition. Note that we were nevertheless able to give an existence result for registration with the mass-preserving transformation operator (compare Theorem 6.3.3).

In order to prove this convergence properties we assumed boundedness of the Banach indicatrix in L^∞. While it is not clear whether this assumption holds for transformations with bounded hyperelastic energy, we shortly sketch out a special case:
For any transformation

$$y \in C^1(\Omega, \mathbb{R}^d) \qquad \det(\nabla y) > 0 \tag{8.5}$$

the implicit function theorem assures us, that for every $x \in \Omega$ there exists an open neighbourhood U_x such that $y|_{U_x}$ is injective. For any compact subset $K \subsetneq \Omega$ we have

$$K \subset \bigcup_{x \in K} U_x. \tag{8.6}$$

Since K is compact the open cover defined by (8.6) has a finite subcover, i.e. there exists a finite set I such that for every $i \in I$ there is $x_i \in K$ with

$$K \subset \bigcup_{i \in I} U_{x_i}. \tag{8.7}$$

Now $y|_{U_{x_i}}$ is injective and therefore the Banach indicatrix is bounded by $\mathrm{card}(I)$. This reasoning is not directly applicable for $y \in H^1$ with $\det(\nabla y) > 0$ a.e, because the transformation is not necessarily invertible at points x with $\det(\nabla y(x)) = 0$ and thus we cannot expect to find an open cover like (8.6). Another question is how we can deal with points $x \in \partial\Omega$, which is the reason why we restricted this presentation to compact sets $K \subsetneq \Omega$.

As an alternative we gave existence results for motion-corrected reconstruction restricted to injective transformations (compare Corollaries 7.2.6 and 7.2.7). Note that the regularity of the cofactors of the admissible transformation were sufficient to guarantee that injectivity is preserved in the limit (Theorem 3.3.8) resp. can be deduced via injective boundary values (Proposition 3.2.9).

Since the analysis we presented for both models relies heavily on the injectivity of the transformation, it is inevitable to design a numerical framework for injective registration. Although Dirichlet boundary conditions should be easier to handle than a general restriction to injective transformations, the determination of the boundary values is subject to further research. Nevertheless using a affine preregistration or recorded movements as these starting guesses, seems like a natural way.

Note that the focus of this thesis lies clearly on the derivation and especially thorough analysis of the combined motion and image estimation models we presented in Chapters 6 and 7. While we tested the resulting methods on artificial data as a proof of concept, an extensive evaluation on clinical data is the next step to go.

9 Appendix

9.1 Computation of Registration Functionals

In order to improve the readability of Section 6.2, we moved the computation of distance measures for Poisson and speckle noise into the appendix.

Poisson Noise

In this section we aim to compute similiarity measures for Poisson noise, for example occurring in emission tomography [125] or microscopy imaging [22]. For the sake of numerical stability we use a shifted version of the Kullback-Leibler data fidelity

$$D(u,f) = u - f\ln(u) + f\ln(f) - f, \qquad \partial_u D(u,f) = 1 - \frac{f}{u}. \tag{9.1}$$

Inserting this into the optimality condition (6.9) we obtain

$$1 - \frac{f_0}{\overline{u}} + T_1^* \left(1 - \frac{f_1}{T_1 \overline{u}}\right). \tag{9.2}$$

Note that the nonlinearity in the equation might cause difficulties for a general analysis. However we can solve (9.2) for the intensity-constant transformation operator as well as for mass-preserving transformations.

Intensity-Constant Transformation Again we start with solving the optimality condition (9.2) for the chosen transformation operator:

$$\begin{aligned} 0 &= 1 - \frac{f_0}{\overline{u}} + \left(T^{ic}\right)^* \left(1 - \frac{f_1}{u(v)}\right) \\ \overline{u}(1 + \det(\nabla v^{-1})) &= f_0 + f_1(v^{-1})\det(\nabla v^{-1}) \\ \overline{u} &= \frac{f_0 + f_1(v^{-1})\det(\nabla v^{-1})}{1 + \det(\nabla v^{-1})} \end{aligned}$$

This is the same map we obtained before for Gaussian noise. Therefore we find again, that

$$\overline{u}(v) = \frac{f_0(v) + f_1 \frac{1}{\det(\nabla v)}}{1 + \frac{1}{\det(\nabla v)}} = \frac{f_0(v)\det(\nabla v) + f_1}{1 + \det(\nabla v)}.$$

With that at hand we start the computation of

$$\begin{aligned} D(\overline{u}, f_0) &= \int_\Omega \overline{u}(x) - f_0(x)\ln(\overline{u}(x)) + f_0(x)\ln(f_0(x)) - f_0(x)\mathrm{d}x \\ &= \int_\Omega (\overline{u}(v(x)) - f_0(v(x))\ln(\overline{u}(v(x))) + f_0(v(x))\ln(f_0(v(x))) - f_0(v(x)))\det(\nabla v)\mathrm{d}x \end{aligned}$$

as well as

$$D(T_{ic}\overline{u}, f_1) = \int_\Omega \overline{u}(v(x)) - f_1(x)\ln(u(v(x))) + f_1(x)\ln(f_1(x)) - f_1(x)\mathrm{d}x.$$

The resulting distance measure is now given by the sum:

$$\begin{aligned} D^{poisson,ic} &= D(\overline{u}, f_0) + D(T_v^{ic}\overline{u}, f_1) \\ &= \int_\Omega (\overline{u}(v(x)) - f_0(v(x))\ln(\overline{u}(v(x))) + f_0(v(x))\ln(f_0(v(x))) - f_0(v(x)))\det(\nabla v) \\ &\qquad + \overline{u}(v(x)) - f_1(x)\ln(u(v(x)) + f_1(x)\ln(f_1(x)) - f_1(x)\mathrm{d}x. \\ &= \int_\Omega \overline{u}(v(x))(1+\det(\nabla v)) - (f_0(v(x))\det(\nabla v) - f_1(x))\ln(u(v(x)) \\ &\qquad + f_0(v(x))\det(\nabla v)(\ln(f_0(v(x))) - 1) + f_1(x)(\ln(f_1(x)) - 1)\mathrm{d}x \\ &= \int_\Omega (f_0(v(x))\det(\nabla v) + f_1(x))\left(1 - \ln\left(\frac{f_0(v(x))\det(\nabla v) + f_1(x)}{1+\det(\nabla v)}\right)\right) \\ &\qquad + f_0(v(x))\det(\nabla v)(\ln(f_0(v(x))) - 1) + f_1(x)(\ln(f_1(x)) - 1)\mathrm{d}x \\ &= \int_\Omega f_0(v(x))\det(\nabla v)\ln(f_0(v)) \\ &\qquad - (f_0(v(x))\det(\nabla v) + f_1(x))\ln\left(\frac{f_0(v(x))\det(\nabla v) + f_1(x)}{1+\det(\nabla v)}\right)\mathrm{d}x \end{aligned}$$

Note that the second row of the next to last conversion is the result of the additive shift $f\ln(f) - f$ of the data fidelity, which cancels some terms in the resulting distance measure.

Mass-Preserving transformation For the mass-preserving transformation we start with the computation of the map by solving (9.2):

$$0 = 1 - \frac{f_0}{\overline{u}} + (T_v^{mp})^*\left(1 - \frac{f_1}{\overline{u}(v)\det(\nabla v)}\right)$$

$$
\begin{aligned}
0 &= 1 - \frac{f_0}{\overline{u}} + 1 - \frac{f_1(v^{-1})\det(\nabla v^{-1})}{\overline{u}} \\
2\overline{u} &= f_0 + f_1(v^{-1})\det(\nabla v^{-1}) \\
\overline{u} &= \frac{f_0 + f_1(v^{-1})\det(\nabla v^{-1})}{2}.
\end{aligned}
$$

Therefore we obtain:

$$
\overline{u}(v) = \frac{f_0(v) + f_1 \frac{1}{\det(\nabla v)}}{2} = \frac{f_0(v)\det(\nabla v) + f_1}{2\det(\nabla v)}.
$$

Inserting this into (6.12) we can derive the effective functional:

$$
\begin{aligned}
D^{poisson,mp} &= D(\overline{u}, f_0) + D(T_v^{mp}\overline{u}, f_1) \\
&= \int_\Omega \overline{u} - f_0 \ln(\overline{u}) + f_0 \ln(f_0) - f_0 \\
&\qquad \overline{u}(v)\det(\nabla v) - f_1 \ln(\overline{u}(v)\det(\nabla v)) + f_1 \ln(f_1) - f_1 \mathrm{d}x \\
&= \int_\Omega (\overline{u}(v) - f_0(v)\ln(\overline{u}(v)) + f_0(v)\ln(f_0(v)) - f_0(v))\det(\nabla v) \\
&\qquad \overline{u}(v)\det(\nabla v) - f_1 \ln(\overline{u}(v)\det(\nabla v)) + f_1 \ln(f_1) - f_1 \mathrm{d}x \\
&= \int_\Omega 2\overline{u}(v)\det(\nabla v) - (f_0(v)\det(\nabla v) + f_1)\ln(\overline{u}(v)\det(\nabla v)) \\
&\qquad + f_0(v)\det(\nabla v)(\ln(f_0(v)\det(\nabla v)) - 1) + f_1(\ln(f_1) - 1)\mathrm{d}x \\
&= \int_\Omega f_0(v)\det(\nabla v)\ln(f_0(v)\det(\nabla v)) + f_1 \ln(f_1) \\
&\qquad - (f_0(v)\det(\nabla v) + f_1)\ln\left(\frac{f_0(v)\det(\nabla v) + f_1}{2}\right)\mathrm{d}x.
\end{aligned}
$$

Note that the non-logarithmic terms in the data fidelity get cancelled out.

Speckle Noise

The data fidelity for speckle noise (occuring for example in ultrasound imaging [1]) is given by

$$
D(f, u) = \frac{(u - f)^2}{u}.
$$

According to our framework we also need the derivative, which is given by

$$
\partial_u D(f, u) = \frac{u^2 - f^2}{u^2}.
$$

With that at hand we can use the optimality condition (6.9) to compute the optimal $\overline{u}$ for our two transformation models. In the registration framework we end up with the following condition:

$$\frac{\overline{u}^2 - f_0^2}{\overline{u}^2} + T_v^* \frac{(T_v\overline{u})^2 - f_1^2}{(T_v\overline{u})^2} = 0. \tag{9.3}$$

Intensity-Constant Transformation We use the given adjoint transformation operator to solve the optimality condition (9.3) for $\hat{u}$:

$$\begin{aligned}
0 &= \frac{\overline{u}^2 - f_0^2}{\overline{u}^2} + \left(T_v^{ic}\right)^* \frac{\overline{u}(v)^2 - f_1^2}{\overline{u}(v)^2} \\
0 &= \frac{\overline{u}^2 - f_0^2}{\overline{u}^2} + \frac{\overline{u}^2 - f_1(v^{-1})^2}{\overline{u}^2} \det(\nabla v^{-1}) \\
0 &= \overline{u}^2 - f_0^2 + (\overline{u}^2 - f_1(v^{-1})^2) \det(\nabla v^{-1}) \\
\overline{u}^2(1 + \det(\nabla v)) &= f_0^2 + f_1(v^{-1})^2 \det{}^{-1}(\nabla v) \\
\overline{u} &= \sqrt{\frac{f_0^2 + f_1(v^{-1})^2 \det(\nabla v^{-1})}{1 + \det(\nabla v^{-1})}}.
\end{aligned}$$

We compute now:

$$\begin{aligned}
D(\overline{u}, f_0) &= \int_\Omega \frac{(\overline{u}(x) - f_0(x))^2}{\overline{u}(x)} \mathrm{d}x \\
&= \int_\Omega \frac{\overline{u}(x)^2 + f_0(x)^2}{\overline{u}(x)} - 2f_0(x) \mathrm{d}x \\
&= \int_\Omega \frac{\overline{u}(v(x))^2 + f_0(v)^2}{\overline{u}(v(x))} \det(\nabla v) - 2f_0(v(x)) \det(\nabla v) \mathrm{d}x. \qquad (9.4) \\
& \qquad (9.5)
\end{aligned}$$

Furthermore we obtain:

$$D(T_v^{ic}u, f_1) = \int_\Omega \frac{\overline{u}(v(x))^2 + f_1(x)^2}{\overline{u}(v(x))} - 2f_1 \mathrm{d}x. \tag{9.6}$$

To carry out the computation of the resulting distance measure we use

$$\overline{u}(v) = \sqrt{\frac{f_0(v)^2 \det(\nabla v) + f_1^2}{1 + \det(\nabla v)}}.$$

Now the resulting distance measure $D^{speckle,ic}$ is given by

$$D^{speckle,ic} = D(\overline{u}, f_0) + D(T_v^{ic}\overline{u}, f_1)$$

$$
\begin{aligned}
&= \int_\Omega \frac{\left(\frac{f_0(v)^2\det(\nabla v)+f_1^2}{1+\det(\nabla v)} + f_0(v)^2\right)\det(\nabla v) + \frac{f_0(v)^2\det(\nabla v)+f_1^2}{1+\det(\nabla v)} + f_1^2}{\overline{u}(v)} \\
&\quad - 2(f_0(v)\det(\nabla v) + f_1)\mathrm{d}x \\
&= \int_\Omega 2\frac{f_0(v)^2\det(\nabla v)+f_1^2}{\sqrt{\frac{f_0(v)^2\det(\nabla v)+f_1^2}{1+\det(\nabla v)}}} - 2(f_0(v)\det(\nabla v)+f_1)\mathrm{d}x \\
&= \int_\Omega 2\sqrt{(f_0(v)^2\det(\nabla v)+f_1^2)(1+\det(\nabla v))} - 2(f_0(v)\det(\nabla v)+f_1)\mathrm{d}x.
\end{aligned}
$$

Mass-Preserving Transformation Again we use the given transformation operator to solve the optimality condition (9.3):

$$
\begin{aligned}
0 &= \frac{\overline{u}^2 - f_0^2}{\overline{u}^2} + (T_v^{mp})^* \frac{(\overline{u}(v)\det(\nabla v))^2 - f_1^2}{(\overline{u}(v)\det(\nabla v))^2} \\
0 &= \frac{\overline{u}^2 - f_0^2}{\overline{u}^2} + \frac{\overline{u}^2\frac{1}{\det(\nabla v^{-1})^2} - f_1(v^{-1})^2}{\overline{u}^2\frac{1}{\det(\nabla v^{-1})^2}} \\
0 &= \frac{\overline{u}^2 - f_0^2}{\overline{u}^2} + \frac{\overline{u}^2 - f_1(v^{-1})^2\det(\nabla v^{-1})^2}{\overline{u}^2} \\
0 &= \overline{u}^2 - f_0^2 + \overline{u}^2 - f_1(v^{-1})^2\det(\nabla v^{-1})^2 \\
2\overline{u}^2 &= f_0^2 + f_1(v^{-1})^2\det(\nabla v^{-1})^2 \\
\overline{u} &= \sqrt{\frac{f_0^2 + f_1(v^{-1})^2\det(\nabla v^{-1})^2}{2}}.
\end{aligned}
$$

We can use this result to compute the data fidelities like in the case of the intensity-constant transformation model,thus we obtain:

$$
\begin{aligned}
D(\overline{u}, f_0) &= \int_\Omega \frac{\overline{u}(v(x))^2 + f_0(v)^2}{\overline{u}(v(x))}\det(\nabla v) - 2f_0(v(x))\det(\nabla v)\mathrm{d}x \\
D(T_v u, f_1) &= \int_\Omega \frac{(\overline{u}(v(x))\det(\nabla v))^2 + f_1(x)^2}{\overline{u}(v(x))\det(\nabla v)} - 2f_1\mathrm{d}x.
\end{aligned}
$$

Adding this two distance terms gives the distance measure for registration:

$$
\begin{aligned}
D^{speckle,mp} &= D(\overline{u}, f_0) + D(T_v^{mp}\overline{(u)}, f_1) \\
&= \int_\Omega \frac{2(\overline{u}(v)\det(\nabla v))^2 + (f_0(v)\det(\nabla v))^2 + f_1^2}{\overline{u}(v)\det(\nabla v)} - 2(f_0(v)\det(\nabla v)+f_1)\mathrm{d}x.
\end{aligned}
\tag{9.7}
$$

As the computation of $\overline{u}(v)\det(\nabla v)$ is crucial we will carry it out:

$$
\overline{u}(v)\det(\nabla v) = \sqrt{\frac{f_0(v)^2 + f_1^2\frac{1}{\det(\nabla v)^2}}{2}}\det(\nabla v)
$$

$$= \sqrt{\frac{(f_0(v)\det(\nabla v))^2 + f_1^2}{2}}.$$

Inserting this into (9.7) we end up with:

$$\begin{aligned} D^{speckle,mp} &= \int_\Omega \frac{2((f_0(v)\det(\nabla v))^2 + f_1^2)}{\sqrt{\frac{(f_0(v)\det(\nabla v))^2+f_1^2}{2}}} - 2(f_0(v)\det(\nabla v) + f_1)\mathrm{d}x \\ &= \int_\Omega 2\sqrt{2((f_0(v)\det(\nabla v))^2 + f_1^2)} - 2(f_0(v)\det(\nabla v) + f_1)\mathrm{d}x. \end{aligned}$$

9.2 Numerical Implementations

For the sake of completeness we present the shortly commented code of the implementations described in the numerical framework for both problems.

9.2.1 Bayesian ML Estimation for Image Registration

Listing 9.1: Objective Function for Registration with noise modeling

```
%Objective Function for registration with Bayesian Modelling. Note that this
%code is matrix free. Furthermore specially designed distance Functions
%have to be used.

function [Jc,para,dJ,H]=BayesianobjFctn(T,Rc,omega,m,~,yc)

Jc = []; para = []; dJ = []; H = [];
persistent P

P = gridInterpolation(P,omega,m);

doDerivative = (nargout>2);                % flag for necessity of derivatives

% compute interpolated image and derivative, formally: center(yc) = P*yc
dim=length(omega)/2;
n   = length(center(yc,m))/dim;
[Tc,dT] = inter(reshape(T,m),omega,center(yc,m),'doDerivative',doDerivative);
```

```
18
19
20  % compute jacobian determinant for intensity modulation
21  [Jac,dJac] = geometry(yc,m,'Jac','doDerivative',doDerivative
        ,'omega',omega);
22
23  % compute distance measure
24  [Dc,rc,dDTc,dDJac,d2psi] = distance(Tc,Rc,omega,m,'
        doDerivative',doDerivative,'Jac',Jac);
25
26  % compute regularizer
27  [Sc,dS,d2S] = regularizer(yc-getNodalGrid(omega,m),omega,m,'
        doDerivative',doDerivative);
28
29  % evaluate joint function and return if no derivatives need
        to be computed
30  Jc = Dc + Sc;
31
32  % store intermediates for outside visualization
33  para = struct('Tc',Tc,'Rc',Rc,'omega',omega,'m',m,'yc',
        center(yc,m),'Jc',Jc);
34
35  if ~doDerivative, return; end;
36
37  %Chain Rule: total derivative results from derivating by Jac
        and Tc:
38
39  dD=P((dDTc*dT)')'+dJac.dJacadj(yc,m,dDJac')';
40
41
42  dJ=dD+dS;
43
44  H=[];
45
46
47
48  end
49
50  function P = gridInterpolation(P,omega,m)
51  switch regularizer
52      case 'mbHyperElastic',
53          if size(P,1) ~= length(omega)/2 * prod(m)
54              P = nodal2center(m);
55          end
```

```
56      case {'mfHyperElastic'},
57          P = @(y) nodal2center(y,m);
58      otherwise
59          error('Bayesian Modelling requires Hyperelastic
               Regularization!');
60  end
61  end
```

This is related to the objective function used in FAIR, but we point out the difference, that the derivative is computed with help of (6.38) (see line 39). Furthermore the distance measures resulting from our noise modeling depend on the Jacobian determinant, thus we need to pass the determinant to the distance measure (line 24). With this two modifications the objective function is capable of performing registration with all the distance measures we derived before. Rather than showing the code for all distances, we focus on the presentation of the distance measure resulting from Gaussian noise with the intensity-constant transformation operator as an example:

Listing 9.2: Intensity-constant Gaussian distance

```
1  %Distance Term for Gaussian noise with intensity-constant
      transformation
2  %operator
3  function [Dc,rc,dDTc,dDJac,d2psi]=Gaussian(Tc,Rc,omega,m,
      varargin)
4  %Initialisation of the output parameters
5  Dc  = []; dDTc = []; rc  = []; dDJac = []; d2psi = [];
6  doDerivative = (nargout > 2);
7  %Noise level parameter, setting default value to 1.
8  sigma=1;
9  Jac=[];
10 for k=1:2:length(varargin), % overwrite default parameter
11   eval([varargin{k},'=varargin{',int2str(k+1),'};']);
12 end;
13
14 Rc=reshape(Rc,[numel(Rc),1]);
15 %Cell size for integration
16 hd  = prod((omega(2:2:end)-omega(1:2:end))./m);
17 %Residual
18 rc=Tc-Rc;
19 %Scaling factor in the distance measure
20 detterm=Jac./(1+Jac);
21 %Value for the distance measure
22 Dc= hd/(2*sigma)*rc'*(rc.*detterm);
23
24 if ~doDerivative, return; end;
```

```
25
26 %Derivative with respect to the transformed template
27 dDTc = hd /(2*sigma) * (2*(Tc-Rc).*detterm)';
28 %Derivative with respect to the Jacobian determinant
29 dDJac=(hd /(2*sigma)* rc.^2.*(1./(Jac+1)-Jac./((Jac+1).^2)))
       ';
30 %2nd derivative is not needed in the Bayesian Framework
31 d2psi = [];
32
33 end
```

As we see this distance measure is implemented similar to the standard distance measures used in the FAIR toolbox, with the difference, that we need the Jacobian determinant as an additional input and assign derivatives with respect to the transformed template and the Jacobian determinant. Having described the implementation of the objective function, we turn our focus on the minimization of this objective function.
As an example we present the implementation for non-parametric image registration (Note that parametric image registration can be implemented in a similar way):

Listing 9.3: Multi-Level Minimization for the Bayesian objective Function

```
1  function [yc] = BayesianMLIR(MLdata,alpha,varargin)
2
3  % setup default parameter for non-parametric registration
4  NPIRopt     = @modlBFGS;          % optimizer to be used for
       NPIR
5  NPIRobj     = @BayesianobjFctn; % objective function for
       NPIR
6  yStop       = [];                 % global stopping for NPIR
7  yRef        = [];                 % regularization: S(yc-yRef)
8  adaptive=1;                       %Flag for adaptive Parameter
       choice
9  ratiov=1;                          %Ratio alphaVolume/
       alphaLength
10 yprereg=[];                       %Possible Starting guess
11 ratioa=0;                        %Ratio alphaArea/Alpha/length
12 [MLdata,minLevel,maxLevel] = getMultilevel(MLdata);
13
14 for k=1:2:length(varargin), % overwrite default parameter
15   eval([varargin{k},'=varargin{',int2str(k+1),'};']);
16 end;
17
18 % initialization
19 dimstr   = @(m) sprintf('[%s]',sprintf(' %d',m));
20 omega    = MLdata{end}.omega;  % spatial domain
```

```
xc      = [];                   % current grid
%Bayesian Frameworks needs hyperelastic Regularization!
regularizer('reset','regularizer','mfHyperElastic','alpha'
   ,1,'alphaLength',1,'alphaArea',alphaArea,'alphaVolume',
   ratio);

%Grid Type
getGrid = @(m) getNodalGrid(omega,m);

fprintf('\n\n');
fprintf('%s: MultiLevel Image Registration\n',mfilename)
fprintf('-- distance=%s, regularizer=%s, alpha=%s, trafo=%s\
   n', ...
  distance,regularizer,num2str(regularizer('get','alpha')),
     trafo);

%For Loop for the Multi-Level Minimization
for level=minLevel:maxLevel,

  %choose adaptive weighting
  dim=length(MLdata{level}.omega)/2;
  basis=2^dim;
  %Chose the regularization weight corresponding to the
     noise variance on the
  %current Level
  gewicht=(basis^(-(maxLevel-level)))^(adaptive);
  regularizer('reset','regularizer','mfHyperElastic','alpha'
     ,gewicht*alpha,'alphaLength',1,'alphaArea',ratioa,'
     alphaVolume',ratiov);

  xOld  = xc;
  m     = MLdata{level}.m;
  xc    = getGrid(m);
  [T,R] = inter('coefficients',MLdata{level}.T,MLdata{level
     }.R,omega);
  Rc    = inter(R,omega,center(xc,m));
  FAIRmessage(sprintf('%s: level %d from %d to %d, %s',...
   mfilename,level,minLevel,maxLevel,dimstr(MLdata{level}.m)
      ));
  NPIRfctn = @(yc) NPIRobj(T,Rc,omega,m,yRef,yc);

if level==minLevel

    yRef=xc;
```

```
56      y0=yRef;
57
58      if ~isempty(yprereg)
59
60          y0=yprereg;
61
62      end
63
64  else
65      % prolongate yc (coarse) y0 (current)
66      y0 = xc + mfPu(yc - xOld,omega,m/2);
67  end;
68  %Solve the MinimizationProblem
69  yc=modlBFGS(NPIRfctn,y0,1,MLdata{level}.omega,MLdata{level}.
        m,1);
70
71  end
```

We outline briefly the main components of this implementation, of our function for Multi-Level image registration in the Bayesian Framework we presented: As we see in the function declaration (line 1), we obtain a transformation from the given Multi-Level data and a regularization parameter.
In the lines 3-12 we just the default parameters for the registration. The values assigned to the parameters are selected more or less arbitrarily based on some tests. Note that all of this hardcoded parameters can be changed by overwriting them in the function call with help of the *varargin* option (line 14-16).
After some technical initializations and some output the backbone of this routine is given by the for loop starting in line 34:
Note that we set an adaptive regularization parameter in the lines 36-42. This adaptive parameter choice links the regularization parameter alpha to the variance of the noise on each level (compare example 5.1.1).
After initializing the objective function (line 44-51) and the starting guess (line 53. 67) in line 69 the actual minimization is performed with the modified BFGS method specified by (5.16) and (5.18).
Having presented the numerical realization of the Bayesian framework for image registration we turn our focus on comparing this framework with existing registration methods in the next section.

9.2.2 Motion Step in Motion-Corrected Reconstruction

The implementation of said function is then given by:

Listing 9.4: Objective Function for the Motion Step

```
%Objective Function for the Motion Step in Motion-Corrected
   Reconstruction
function [Jc,para,dJ,H]=NPIRprojectedobjfctn(T,Rc,omega,m,~,
   yc)
Jc = []; para = []; dJ = []; H = [];
persistent P
datares=[192,192,175];
hd  = prod((omega(2:2:end)-omega(1:2:end))./m);
doDerivative = (nargout>2);

% operator for nodal-->cell centered projections
P = @(y) nodal2center(y,m);

% compute jacobian determinant for intensity modulation
[Jac,dJac] = geometry(yc,m,'Jac','doDerivative',doDerivative
   ,'omega',omega);
%We only use matrix free implementation
matrixFree=1;
% interpolation and apply intensity modulation
[Tc,dT] = inter(reshape(T,m),omega,center(yc,m),'
   doDerivative',doDerivative,'matrixFree',matrixFree);
Tcmod = Tc .* Jac;
%Project the transformed image into the detector domain
Tproj = operator(Tcmod,m,datares);


% Evaluate the Distance Term
[Dc,~,dD,dres,drestrans,d2psi] = distance(Tproj,Rc,omega,
   datares,'doDerivative',doDerivative);

%Compute the regularization Energy
[Sc,dS,d2S] = regularizer(yc-getNodalGrid(omega,m),omega,m,'
   doDerivative',doDerivative);


Jc=Dc+Sc;
if ~doDerivative, return; end


```

```
34
35 %Compute the part of the derivative resulting from the
      projection
36 dD=adjointoperator(dD',m,datares)';
37
38 %Compute the Derivative for the distance measure like in the
       VAMPIRE toolbox
39 dD = (P(vecXmat(dD',Jac,dT)) + dJac.dJacadj(yc,m,dD'.*Tc))';
40
41 %Derivative for the joint functional
42 dJ = dD + dS;
43
44
45 end
46
47
48 function data=operator(image,imres,datares)
49 %Function for the Projection Operator
50 %Here we use the projection operator provided by EMRecon
51 image=reshape(image,imres);
52 writeimage(image,'image');
53 !export OMP_NUM_THREADS=4;echo "4" | ./
      EMrecon_siemens_pet_tools parmfile.emrecon image data >
      bruellfile
54 data = loadimage('data',datares(1),datares(2),datares(3),1);
55 data=reshape(data,[numel(data),1]);
56 end
57
58
59 function image=adjointoperator(data,imres,datares)
60 %Function for the adjoint operator
61 %Again we use the operator provided by EMRecon
62
63 data=reshape(data,datares);
64 writeimage(data,'data');
65 !export OMP_NUM_THREADS=4;echo "5" | ./
      EMrecon_siemens_pet_tools parmfile.emrecon data image >
      bruellfile
66 image = loadimage('image',imres(1),imres(2),imres(3),1);
67 image=reshape(image,[numel(image),1]);
68
69 end
70
71
```

```
72
73  function p = vecXmat(dD,Jac,dT)
74  % implementation of vector' * matrix product dD * [Jac] * dT
75  % where
76  %
77  % size(dD)  = [1, nVol * prod(m)]
78  % size(Jac) = [prod(m),1]
79  % size(dI)  = [prod(m),dim,nVol]
80  dim = size(dT,2);
81  p   = zeros(length(Jac),dim);
82  for d=1:dim,
83      p(:,d) = dD(:).*Jac(:).*dT(:,d);
84  end
85  p = reshape(p,1,[]);
86  end
87
88  function a = sdiag(a)
89  % shortcut for sparse diagonal matrices
90  a = spdiags(reshape(a,[],1),0,length(a),length(a));
91  end
92
93
94  function p= vecmatprod1(Jac,dI,x)
95  % We aim to compute diag(Jac) * [dI] * x
96  %
97  % size(Jac) = [n, 1]
98  % size(dI) = [nVol * n , dim * n]
99  % size(x)  = [dim*n,1]
100 dim  = size(dI,2);
101 nVol = size(dI,3);
102 n    = length(Jac);
103
104 x = reshape(x,[],dim);
105 p = zeros(n,nVol);
106 for vol=1:nVol,
107     for d=1:dim,
108         p(:,vol) = p(:,vol) + dI(:,d,vol) .* x(:,d);
109     end
110     p(:,vol) = p(:,vol).*Jac;
111 end
112 p = p(:);
113 end
114
115 function p = vecmatprod1adj(dI,Jac,x)
```

```
116 % We aim to compute dI' * [Jac] * x
117 %
118 % size(Jac) = [n,1]
119 % size(dI)  = [nVol*n, dim*n]
120 % size(x)   = [nVol*n,1]
121 dim  = size(dI,2);
122 nVol = size(dI,3);
123 n = length(Jac);
124
125 x = reshape(x,[],nVol);
126 p = zeros(n,dim);
127 for vol=1:nVol,
128     for d=1:dim,
129         p(:,d) = p(:,d) + dI(:,d,vol) .* x(:,vol);
130     end
131 end
132 for d=1:dim,
133     p(:,d) = p(:,d).*Jac;
134 end
135 p = p(:);
136 end
```

As we see this objective function is quite similar to the objective function for mass-preserving image registration provided by the VAMPIRE toolbox ([63]). The difference lies in projecting the transformed template image into the detector domain (line 20), which results in different derivatives (lines 35-36). The projection operator can be chosen in the nested functions (lines 48-69). Note that the operator as well as the adjoint needs to implemented. The nested functions in line 73-136 are provided by the VAMPIRE toolbox and needed to handle the Jacobian determinant.
With this objective function at hand, we can implement arbitrary distance measures. As a example we present the Kullback-Leibler divergence, which we use for motion-corrected PET reconstruction:

Listing 9.5: Kullback-Leibler divergence as Distance Measure

```
1 function [Dc,rc,dD,dr,drtrans,d2psi] = KullbackLeibler(Tc,Rc
     ,omega,datares,varargin)
2
3 Dc  = []; dD = []; rc  = []; dr = []; d2psi = []; drtrans
     =[];
4
5
6 %Treshold for the Kullback-Leibler divergence
7 g=1e-8;
8
```

```
for k=1:2:length(varargin), % overwrite default parameter
  eval([varargin{k},'=varargin{',int2str(k+1),'};']);
end;


rc=Tc-Rc;
%Compute the Distance, where 1/prod(datares) is the
   integration weight for
%each point in the detector domain
Dc=1/prod(datares)*sum(rc-Rc.*log(max(rc+Rc,g))+Rc.*log(max(
   Rc,g)));



if nargout <3, return; end;

%Derivative of the Distance measure
dD=1/prod(datares)*(1-Rc./max(rc+Rc,g))';

%No approximation for the Hessian is needed
d2psi=[];



end
```

A short Comment on Notation

As a consistent notation is crucial for keeping the track in all presented methods we will give a short overview about the entities and the nomenclature we use:

Entity	Nomenclature
Dimension of a $\mathbb{R}$-Vectorspace	d; m
Image Domain	$\Omega \subset \mathbb{R}^d$
Not specified image/function	$u : \Omega \to \mathbb{R}$
Density Image	small greek letters, f.e. ρ
Not specified Hilbert/Banach spaces	big callicraphic latters, f.e. $\mathcal{X}$
transformations	roman letters y, v
(compact) linear operators	big roman letters K, T
Functional	big roman letter J
functions	small roman letters j, f, g

Table 9.1: The notation used throughout this thesis

List of Figures

List of Tables

Bibliography

[1] A. Achim, A. Bezerianos, and P. Tsakalides. Novel bayesian multiscale method for speckle removal in medical ultrasound images. *Medical Imaging, IEEE Transactions on*, 20(8):772–783, 2001. 159

[2] R.A. Adams and J.J.F. Fournier. *Sobolev Spaces*. Pure and Applied Mathematics. Elsevier Science, 2003. 13, 15

[3] S. Allassonnière, Y. Amit, and A. Trouvé. Towards a Coherent Statistical Framework for Dense Deformable Template Estimation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 69(1):pp. 3–29, 2007. 8, 99

[4] H. W. Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. Springer Lehrbuch. Springer, 2002. 30, 122

[5] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, Oxford, New York, 2000. 18, 47, 48

[6] J. Ashburner. A Fast Diffeomorphic Image Registration Algorithm. *NeuroImage*, 38(1):95–113, 2007. 146

[7] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Lojasiewicz Inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010. 154

[8] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV Spaces*. MPS-SIAM Series on Optimization, 2006. 46, 47

[9] A. B. Bakushinskii. Remarks on Choosing a Regularization Parameter Using the Quasi-Optimality and Ratio Criterion. *USSR Computational Mathematics and Mathematical Physics*, 24(4):181 – 182, 1984. 36

[10] S. Banach. Sur les lignes rectifiables et les surfaces dont l'aire est finie. *Fundamenta Mathematicae*, 7(1):225–236, 1925. 21

[11] S. Barendt and J. Modersitzki. A Variational Model for SPECT Reconstruction with a Nonlinearly Transformed Attenuation Prototype. *International Journal of Computer Mathematics*, 90(1):82–91, 2013. 150

[12] H. Bauer. *Probability Theory*. Bod Third Party Titles, 1996. 39

[13] H. Bauer. *Measure and Integration Theory*. W. de Gruyter, 2001. 13

[14] T. Bayes and R. Price. An Essay towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS. *Philosophical Transactions (1683-1775)*, pages 370–418, 1763. 40

[15] A. Beck. On the Convergence of Alternating Minimization for Convex Programming with Applications to Iteratively Reweighted Least Squares and Decomposition Schemes. *SIAM Journal on Optimization*, 25(1):185–209, 2015. 153

[16] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. Taylor & Francis, 1998. 34

[17] M. Blume, A. Martinez-Möller, A. Keil, N. Navab, and M. Rafecas. Joint Reconstruction of Image and Motion in Gated Positron Emission Tomography. *IEEE Transactions on Medical Imaging*, 29(11):1892–1906, November 2010. 147

[18] M. Blume, N. Navab, and M. Rafecas. Joint Image and Motion Reconstruction for PET using a B-Spline Motion Model. *Phys Med Biol*, 57(24):8249–8270, Dec 2012. 117, 147

[19] K. Bredies, K. Kunisch, and T. Pock. Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492 – 526, 2010. 48

[20] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer, 2010. 14

[21] L.E.J. Brouwer. Über Abbildung von Mannigfaltigkeiten. *Mathematische Annalen*, 71(1):97–115, 1911. 23

[22] C. Brune, A. Sawatzky, and M. Burger. Bregman-EM-TV Methods with Application to Optical Nanoscopy. In Xue-Cheng Tai, Knut Märken, Marius Lysaker, and Knut-Andreas Lie, editors, *Scale Space and Variational Methods in Computer Vision*, volume 5567 of *Lecture Notes in Computer Science*, pages 235–246. Springer Berlin / Heidelberg, 2009. 2, 117, 134, 135, 138, 157

[23] A. Buades, B. Coll, and J. M. Morel. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005. 7

[24] R. Burden and J. Faires. *Numerical Analysis*. Cengage Learning, 2010. 58, 59

[25] M. Burger, J. Modersitzki, and L. Ruthotto. A Hyperelastic Regularization Energy for Image Registration. *SIAM Journal on Scientific Computing*, 35(1):B132–B148, 2013. 47, 53, 54

[26] M. Burger and S. Osher. A Guide to the TV Zoo. In *Level Set and PDE Based Reconstruction Methods in Imaging*, Lecture Notes in Mathematics, pages 1–70. Springer International Publishing, 2013. 46, 47, 48, 125

[27] M. Burger and O. Scherzer. Regularization Methods for Blind Deconvolution and Blind Source Separation Problems. *Math. Cont. Signals & Systems*, 2001, 2001. 153

[28] F. Büther, M. Dawood, L. Stegger, F. Wübbeling, M. Schäfers, O. Schober, and K. P. Schäfers. List Mode-Driven Cardiac and Respiratory Gating in PET. *Journal of Nuclear Medicine*, 50(5):674–681, 2009. 1, 2

[29] M. Carlavan and L. Blanc-Féraud. Sparse Poisson Noisy Image Deblurring. *IEEE Transactions on Image Processing*, 21(4):1834–1846, 2012. 2, 137

[30] P. Celada, G. Cupini, and M. Guidorzi. A Sharp Attainment Result for Nonconvex Variational Problems. *Calculus of Variations and Partial Differential Equations*, 20(3):301–328, 2004. 30, 111

[31] P. Celada and S. Perrotta. On the Minimum Problem for Nonconvex, Multiple Integrals of Product Type. *Calculus of Variations and Partial Differential Equations*, 12(4):371–398, 2001. 111

[32] A. Chambolle and T. Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. 154

[33] R. H. Chan and K. Chen. Multilevel Algorithm for a Poisson Noise Removal Model with Total-Variation Regularization. *Int. J. Comput. Math.*, 84(8):1183–1198, August 2007. 105

[34] T. Chan, A. Marquina, and P. Mulet. High-Order Total Variation-Based Image Restoration. *SIAM Journal on Scientific Computing*, 22(2):503 – 516, 2000. 48

[35] M. Cheney, D. Isaacson, and J. C. Newell. Electrical impedance tomography. *SIAM review*, 41(1):85–101, 1999. 42

[36] P.G. Ciarlet. *Mathematical Elasticity: Three-Dimensional Elasticity*. Number Bd. 1 in Mathematical Elasticity: Three-dimensional Elasticity. North-Holland, 1993. 53

[37] J. B. Conway. *A Course in Functional Analysis*. Graduate Texts in Mathematics. World Publishing Company, 1990. 29, 30, 31, 32

[38] A. M. Cormack. Representation of a Function by its Line Integrals, with some Radiological Applications. *Journal of Applied Physics*, 34(9):2722 – 2727, 1963. 1, 43

[39] B. Dacorogna. *Introduction to the Calculus of Variations*. Imperial College Press, 2004. 28

[40] M. Defrise and P. Kinahan. Data Acquisition and Image Reconstruction for 3D PET. In Bernard Bendriem and DavidW. Townsend, editors, *The Theory and Practice of 3D PET*, volume 32 of *Developments in Nuclear Medicine*, pages 11–53. Springer Netherlands, 1998. 2, 42

[41] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge. Joint Reconstruction of PET-MRI by Exploiting Structural Similarity. *Inverse Problems*, 31(1):015001, 2015. 152

[42] H. Engels. *Numerical Quadrature and Cubature*. Academic Press, 1980. 58

[43] H. W. Engl. Discrepancy principles for Tikhonov Regularization of Ill-Posed Problems Leading to Optimal Convergence Rates. *Journal of Optimization Theory and Applications*, 52(2):209–215, 1987. 37

[44] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer, 1996. 35, 36, 38, 133

[45] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 1998. 15, 16, 17, 67, 68, 77, 123

[46] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. Taylor & Francis, 1991. 19

[47] H. Federer. *Geometric Measure Theory*. Grundlehren der mathematischen Wissenschaften. Springer, 1969. 19, 26

[48] B. Fischer and J. Modersitzki. Ill-posed Medicine - An Introduction to Image Registration. *Inverse Problems*, 24:1–19, 2008. 49

[49] I. Fonseca and W. Gangbo. Local Invertibility of Sobolev Functions. *SIAM Journal on Mathematical Analysis*, 26(2):280–304, 1995. 54

[50] M. Fréchet. Sur la notion de différentielle. *C.R. Acad. Sci. Paris*, 152:845–847;1050–1051, 1911. 66

[51] M. Frechet. Sur la notion de différentielle totale. In *Nouvelles Annales de mathématiques*, volume 12, pages 385–403. Gauthier-Villars, 1912. 66

[52] K. Frick, P. Marnitz, and A. Munk. Statistical Multiresolution Estimation for Variational Imaging: With an Application in Poisson-Biophotonics. *Journal of Mathematical Imaging and Vision*, 46(3):370–387, 2013. 105, 106

[53] R. Gateaux. Sur les fonctionnelles continues et les fonctionnelles analytiques. *CR Acad. Sci. Paris*, 157:325–327, 1913. 65

[54] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741, November 1984. 41

[55] M. Giaquinta and G. Modica. *Mathematical Analysis: An Introduction to Functions of Several Variables*. Mathematical analysis. Birkhäuser Boston, 2009. 18

[56] M. Giaquinta, G. Modica, and J. Soucek. Cartesian Currents, Weak Diffeomorphisms and Existence Theorems in Nonlinear Elasticity. *Archive for Rational Mechanics and Analysis*, 106(2):97–159, 1989. 24, 27

[57] M. Giaquinta, G. Modica, and J. Soucek. Area and the Area Formula. *Rendiconti del Seminario Matematico e Fisico di Milano*, 62(1):53–87, 1992. 18

[58] M. Giaquinta, G. Modica, and J. Soucek. A Weak Approach to Finite Elasticity. *Calculus of Variations and Partial Differential Equations*, 2(1):65–100, 1994. 23, 24, 25, 26, 27

[59] M. Giaquinta, G. Modica, and J. Soucek. Remarks on the Degree Theory. *Journal of Functional Analysis*, 125(1):172 – 200, 1994. 23, 27, 55

[60] M. Giaquinta, G. Modica, and J. Soucek. *Cartesian Currents in the Calculus of Variations I: Cartesian Currents*. A Series of Modern Surveys in Mathematics Series. Springer, 1998. 18, 22, 24, 26, 27, 120, 124, 147, 148

[61] M. Giaquinta, G. Modica, and J. Soucek. Some Regularity Properties of Locally Weakly Invertible Maps. In Giuseppe Buttazzo, GiovanniPaolo Galdi, Ermanno Lanconelli, and Patrizia Pucci, editors, *Nonlinear Analysis and Continuum Mechanics*, pages 45–59. Springer New York, 1998. 27, 55

[62] M. Giaquinta, G. Modica, and J. Soucek. *Cartesian Currents in the Calculus of Variations II: Variational Integrals*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer Berlin Heidelberg, 2010. 24, 27

[63] F. Gigengack, L. Ruthotto, M. Burger, C.H. Wolters, X. Jiang, and K.P. Schafers. Motion Correction in Dual Gated Cardiac PET Using Mass-Preserving Image Registration. *IEEE Transactions on Medical Imaging*, 31(3):698–712, March 2012. 2, 87, 117, 171

[64] T. Gill, S. Basu, W. Zachary, and V. Steadman. Adjoint for Operators in Banach Spaces. *Proceedings of the American Mathematical Society*, 132(5):1429–1434, 2004. 31

[65] E. Haber. *Computational Methods in Geophysical Electromagnetics:* . Mathematics in Industry. Society for Industrial and Applied Mathematics, 2014. 57

[66] E. Haber and J. Modersitzki. Intensity Gradient Based Registration and Fusion of Multi-Modal Images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 726–733. Springer, 2006. 51

[67] W. Hackbusch. Multi-Grid Convergence Theory. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, volume 960 of *Lecture Notes in Mathematics*, pages 177–219. Springer Berlin Heidelberg, 1982. 63

[68] J. Hadamard. Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902. 34

[69] P. Hajłasz. Change of Variables Formula under Minimal Assumptions. *Colloq. Math.*, 64(1):93–101, 1993. 18, 19, 21, 22

[70] P. C. Hansen. The Truncated SVD as a Method for Regularization. *BIT Numerical Mathematics*, 27(4):534–553, 1987. 37

[71] P. C. Hansen. Regularization Tools: A Matlab Package for Analysis and Solution of Discrete Ill-Posed Problems. *Numerical Algorithms*, 6(1):1–35, 1994. 36, 43, 44, 175

[72] P. C. Hansen. Regularization Tools version 4.0 for Matlab 7.3. *Numerical Algorithms*, 46(2):189–194, 2007. 36, 43

[73] P. C. Hansen, T. Sekii, and H. Shibahashi. The Modified Truncated SVD Method for Regularization in General Form. *SIAM Journal on Scientific and Statistical Computing*, 13(5):1142–1150, 1992. 37

[74] M. Hazewinkel. *Encyclopaedia of Mathematics (set)*. Springer Netherlands. 20

[75] D. Henao and C. Mora-Corral. Invertibility and Weak Continuity of the Determinant for the Modelling of Cavitation and Fracture in Nonlinear Elasticity. *Arch. Ration. Mech. Anal.*, 197(2):619–655, 2010. 27

[76] M. Hinze. Discrete Concepts in PDE Constrained Optimization. In *Optimization with PDE Constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*, pages 157–232. Springer Netherlands, 2009. 69

[77] M. Hinze and F. Tröltzsch. Discrete Concepts versus Error Analysis in PDE-Constrained Optimization. *GAMM-Mitteilungen*, 33(2):148–162, 2010. 69

[78] G. N. Hounsfield. Computerized Transverse Axial Scanning (Tomography): Part 1. Description of System. *The British Journal of Radiology*, 46(552):1016–1022, 1973. 1, 43

[79] J. Idier. *Bayesian Approach to Inverse Problems*. ISTE. Wiley, 2013. 39, 41

[80] M. S. Judenhofer and S. R. Cherry. Applications for Preclinical PET/MRI . *Seminars in Nuclear Medicine*, 43(1):19 – 29, 2013. PET/MRI. 152

[81] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160. Springer, 2005. 37, 38, 39, 41

[82] J. B. Keller. Inverse Problems. *The American Mathematical Monthly*, 83(2):pp. 107–118, 1976. 34

[83] D. Kinderlehrer and P. Pedregal. Characterizations of Young Measures Generated by Gradients. *Archive for Rational Mechanics and Analysis*, 115(4):329–365, 1991. 30

[84] K.P. Köesters, T.and Schäfers and F. Wübbeling. EMrecon: An Expectation Maximization Based Image Reconstruction Framework for Emission Tomography Data. *In NSS/MIC Conference Record, IEEE*, pages 4365–4368, 2011. 12, 142

[85] T. Koesters, M. Fieseler, T. Block, F. Wübbeling, D. Faul, F. Boada, and K. P. Schafers. Combined AW-OSEM Reconstruction and Mass-Preserving Motion Correction of PET Data. In *2th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2013. 145, 150

[86] K. Koumatos, F. Rindler, and E. Wiedemann. Orientation-Preserving Young Measures. *arXiv preprint arXiv:1307.1007*, 2013. 111

[87] L. V. Kovalev, J. Onninen, and K. Rajala. Invertibility of Sobolev Mappings under Minimal Hypotheses. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, 27(2):517 – 528, 2010. 54

[88] A. S. Kronrod. On Functions of Two Variables. *Uspehi Matem. Nauk (N.S.)*, 5(1(35)):24–134, 1950. 21

[89] G. Landi and E. Loli Piccolomini. An Improved Newton Projection Method for Nonnegative Deblurring of Poisson-Corrupted Images with Tikhonov Regularization. *Numerical Algorithms*, 60(1):169–188, 2012. 2, 137

[90] D.-H. Li and M. Fukushima. A Modified BFGS Method and its Global Convergence in Nonconvex Minimization. *Journal of Computational and Applied Mathematics*, 129:15–35, 2001. 62

[91] D.-H. Li and M. Fukushima. On the Global Convergence of the BFGS Method for Nonconvex Unconstrained Optimization Problems. *SIAM Journal on Optimization*, 11(4):1054–1064, 2001. 102, 103

[92] D. C. Liu and J. Nocedal. On the Limited Memory BFGS Method for Large Scale Optimization. *Math. Program.*, 45(3):503–528, December 1989. 63

[93] S. M. Lozinskiĭ. On the Indicatrix of Banach. *Vestnik Leningrad. Univ.*, 13(7):70–87, 1958. 21

[94] Nikolai Nikolaevich Luzin. *The Integral and Trigonometric Series*. PhD thesis, Moscow-Leningrad, 1915. 20

[95] J. B. A. Maintz and M. A. Viergever. A Survey of Medical Image Registration. *Medical image analysis*, 2(1):1–36, 1998. 6

[96] B. A. Mair, D. R. Gilland, and J. Sun. Estimation of Images and Nonrigid Deformations in Gated Emission CT. *IEEE Transactions on Medical Imaging*, 25:1130–1144, 2006. 2, 86, 117, 147

[97] D. Mannweiler, J. Rasch, S. Schmid, F. Wübbeling, and K. P. Schäfers. PET Reconstruction of Freely-Moving Mice Comprising Registration-Based Motion Correction of Internal Structures. In *IEEE Nuclear Science Symposium & Medical Imaging Conference*, 2014. 152

[98] J. Martens. Deep Learning via Hessian-Free Optimization. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 735–742, 2010. 62

[99] S. Müller, T. Qi, and B. S. Yan. On a New Class of Elastic Deformations not Allowing for Cavitation. *Annales de l'institut Henri Poincaré (C) Analyse non linéaire*, 11(2):217–243, 1994. 55

[100] J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, 2004. 10, 13, 49, 53, 103, 150

[101] J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration*. SIAM, 2009. 11, 12, 49, 51, 52, 53, 57, 58, 59, 63, 69, 83, 85, 87, 135, 175

[102] V. A. Morozov. On the Solution of Functional Equations by the Method of Regularization. *Soviet Math. Dokl.*, 7:414–417, 1966. 36

[103] K. W. Morton and D. F. Mayers. *Numerical Solution of Partial Differential Equations: An Introduction*. Cambridge University Press, New York, NY, USA, 2005. 69

[104] J.R. Munkres. *Analysis On Manifolds*. Advanced Books Classics Series. Westview Press, 1997. 17, 18

[105] J. Nakamura. *Image Sensors and Signal Processing for Digital Still Cameras*. Optical Science and Engineering. CRC Press, 2005. 73

[106] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, 2001. 42

[107] J. Necas and J. Malek. *Weak and Measure-Valued Solutions to Evolutionary PDEs*. Applied Mathematics. Taylor & Francis, 1996. 30

[108] U. Niesen, D. Shah, and G. W. Wornell. Adaptive Alternating Minimization Algorithms. *IEEE Transactions on Information Theory*, 55(3):1423–1429, March 2009. 153

[109] J. Nocedal. Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation*, 35(151):773–782, 1980. 63

[110] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999. 59, 60, 61, 62, 180

[111] P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock. An Iteratively Reweighted Algorithm for Non-Smooth Non-Convex Optimization in Computer Vision. 2014. Technical Report. 154

[112] F. P. M. Oliveira and J. M. R. S. Tavares. Medical Image Registration: a Review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93, 2014. 6

[113] N. Papenberg. *Ein genereller Registrierungsansatz*. PhD thesis, Universität zu Lübeck, 2008. 57, 58, 175

[114] A. Perry. A Class of Conjugate Gradient Algorithms with a Two-Step Variable Metric Memory. Discussion Papers 269, Northwestern University, Center for Mathematical Studies in Economics and Management Science, January 1977. 63

[115] D. L. Phillips. A Technique for the Numerical Solution of Certain Integral Equations of the First Kind. *J. Assoc. Comput. Mach.*, 9:84–97, 1962. 38

[116] A. Potschka. *A Direct Method for Parabolic PDE Constrained Optimization Problems*. Advances in Numerical Mathematics. Springer Vieweg. in Springer Fachmedien Wiesbaden GmbH, 2013. 69

[117] F. Qiao, T. Pan, J. W. Clark, and O. R. Mawlawi. A Motion-Incorporated Reconstruction Method for Gated PET Studies. *Phys Med Biol*, 51(15):3769–3783, Aug 2006. 117, 152

[118] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Akad. Wiss.*, 69:262–277, 1917. 1, 43

[119] M. Reed and B. Simon. *I: Functional Analysis*. Methods of Modern Mathematical Physics. Elsevier Science, 1981. 15, 28

[120] L. Reichel and G. Rodriguez. Old and new Parameter Choice Rules for Discrete Ill-Posed Problems. *Numerical Algorithms*, 63(1):65–87, 2013. 37

[121] L. Reichel and Q. Ye. Simple Square Smoothing Regularization Operators. *Electron. Trans. Numer. Anal*, 33:63–83, 2009. 38

[122] E. Resmerita and R.S. Anderssen. Joint Additive Kullback-Leibler Residual Minimization and Regularization for Linear Inverse Problems. *Math. Meth. Appl. Sci.*, 30:1527–1544, 2007. 120, 124

[123] A. Roche, G. Malandain, and N. Ayache. Unifying Maximum Likelihood Approaches in Medical Image Registration. *International Journal of Imaging Systems and Technology*, 11(1):71–80, 2000. 71, 99

[124] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In *Medical Image Computing and Computer-Assisted Interventation MICCAI 98*, pages 1115–1124. Springer, 1998. 51

[125] I. Rodrigues, J. Sanches, and J. Bioucas-Dias. Denoising of medical images corrupted by poisson noise. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 1756–1759, Oct 2008. 118, 157

[126] W.C. Röntgen, A. Haase, G. Landwehr, and E. Umbach. *Röntgen Centennial: X-Rays in Natural and Life Sciences*. World Scientific, 1997. 1, 174

[127] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Phys. D*, 60(1-4):259–268, 1992. 46

[128] L. Ruthotto. Mass-Preserving Registration of Medical Images. Master's thesis, WWU Münster, 2010. 50

[129] L. Ruthotto. *Hyperelastic Image Registration*. PhD thesis, WWU Münster, 2012. 54, 55, 103, 119, 127, 129

[130] L. Ruthotto, F. Gigengack, M. Burger, C.H. Wolters, X. Jiang, K. Schäfers, and J. Modersitzki. A Simplified Pipeline for Motion Correction in Dual Gated Cardiac PET. In *Bildverarbeitung für die Medizin 2012*. Springer, 2012. 128

[131] A. Sawatzky, C. Brune, T. Kösters, F. Wübbeling, and M. Burger. EM-TV Methods for Inverse Problems with Poisson Noise. In *Level Set and PDE Based Reconstruction Methods in Imaging*, volume 2090 of *Lecture Notes in Mathematics*, pages 71–142. Springer International Publishing, 2013. 2, 153

[132] A. Sawatzky, C. Brune, F. Wübbeling, T. Kösters, K. P. Schafers, and M. Burger. Accurate EM-TV Algorithm in PET with Low SNR. In *Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE*, pages 5133–5137, Oct 2008. 2

[133] K. P. Schäfers, B. Könemann, B. Czekalla, K. Bolwin, F. Büther, M. Fieseler, H. Braun, S. Ziegler, and H. Quick. Human Thorax Phantom for Simulation of Respiratory and Ccardiac Motion in PET/MRI: Development and First Measurements. In *NSS/MIC Conference Record, IEEE*, 2011. 144

[134] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui. 4D XCAT Phantom for Multimodality Imaging Research. *Medical physics*, 37(9):4902–4915, 2010. 97, 98, 142, 176

[135] M. Sermesant, C. Forest, X. Pennec, H. Delingette, and N. Ayache. Deformable Biomechanical Models: Application to 4D Cardiac Image Analysis. *Medical Image Analysis*, 7(4):475 – 488, 2003. Medical Image Computing and Computer Assisted Intervention. 99

[136] D. F. Shanno. On the Convergence of a New Conjugate Gradient Algorithm. *SIAM Journal on Numerical Analysis*, 15(6):pp. 1247–1257, 1978. 63

[137] L. A. Shepp and Y. Vardi. Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Transactions on Medical Imaging*, 1:113–122, 1982. 2, 117

[138] J. Spijkerman, D. Fontanarosa, M. Das, and W. van Elmpt. Validation of Nonrigid Registration in Pretreatment and Follow-up PET/CT Scans for Quantification of Tumor Residue in Lung Cancer Patients. *Journal of Applied Clinical Medical Physics*, 15(4), 2014. 6

[139] L. A. Steen and J. A. Seebach. *Counterexamples in Topology*. Dover Books on Mathematics Series. Dover Publications, 1995. 20

[140] A. M. Stuart. Inverse Problems: A Bayesian Perspective. *Acta Numerica*, 19:451–559, 5 2010. 41

[141] K. Stüben and U. Trottenberg. Multigrid Methods: Fundamental Algorithms, Model Problem Analysis and Applications. In W. Hackbusch and U. Trottenberg, editors, *Multigrid Methods*, volume 960 of *Lecture Notes in Mathematics*, pages 1–176. Springer Berlin Heidelberg, 1982. 63

[142] S. Suhr, D. Tenbrinck, M. Burger, and J. Modersitzki. Registration of Noisy Images via Maximum A-Posteriori Estimation. In Sébastien Ourselin and Marc Modat, editors, *Biomedical Image Registration*, volume 8545 of *Lecture Notes in Computer Science*, pages 231–240. Springer International Publishing, 2014. 86, 87, 97

[143] A. N. Tikhonov. On the Stability of Inverse Problems. In *Dokl. Akad. Nauk SSSR*, volume 39, pages 195–198, 1943. 38

[144] A. N. Tikhonov. Solution of Incorrectly Formulated Problems and the Regularization Method. *Soviet Math. Dokl.*, 5:1035/1038, 1963. 38

[145] A.G. Vitushkin. On Higher-Dimensional Variations. Master's thesis, Moscow, 1955. 21

[146] D. Werner. *Funktionalanalysis*. Springer-Lehrbuch. Springer London, Limited, 2007. 30, 101

[147] H. Whitney. On Totally Differentiable and Smooth Functions. In James Eells and Domingo Toledo, editors, *Hassler Whitney Collected Papers*, Contemporary Mathematicians, pages 353–369. Birkhäuser Boston, 1992. 19

[148] L. C. Young. Generalized Curves and the Existence of an Attained Absolute Minimum in the Calculus of Variations. *C. R. Soc. Sci. Varsovie, Cl. III*, 30:212–234, 1937. 30

[149] W. Zhou and L. Zhang. Global Convergence of the Nonmonotone {MBFGS} Method for Nonconvex Unconstrained Minimization. *Journal of Computational and Applied Mathematics*, 223(1):40 – 47, 2009. 103

[150] A. Zhu, D. Lee, and H. Shim. Metabolic Positron Emission Tomography Imaging in Cancer Detection and Therapy Response. In *Seminars in Oncology*, volume 38, pages 55–69. Elsevier, 2011. 1

[151] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The Finite Element Method: Its Basis and Fundamentals: Its Basis and Fundamentals*. Elsevier Science, 2005. 69