Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Theoretical and Numerical Insights into Simultaneous Motion and Image Reconstruction for Dynamic Magnetic Particle Imaging

Dissertation with the aim of achieving a doctoral degree at the Faculty of Mathematics, Informatics and Natural Sciences

Department of Mathematics
University of Hamburg

submitted by

Lena Alisha Westen (née Zdun)

2024 in Hamburg

Accepted as dissertation by the Department of Mathematics of the University of Hamburg based on the evaluations by:

Prof. Dr. Christina Brandt (University of Greifswald)
Prof. Dr. Martin Burger (Deutsches Elektronen-Synchrotron DESY and University of Hamburg)

Date of the oral defense: March 27th, 2025

Danksagung

An dieser Stelle möchte ich mich bei all denen bedanken, die mich während der Promotionszeit und beim Anfertigen meiner Dissertation unterstützt haben.

Mein besonderer Dank gilt zunächst Christina Brandt. Du hast mich nicht nur ermutigt, den Weg der Promotion einzuschlagen, sondern warst mir über all die Jahre hinweg eine stetige fachliche und persönliche Stütze. Ohne deine Unterstützung hätte ich diesen Weg wohl nicht begonnen – und sicher nicht so konsequent fortgeführt.
Ebenso danke ich Martin Burger herzlich dafür, sich bereit erklärt zu haben, meine Dissertation zu begutachten.
Ein großes Dankeschön geht an Tobias Kluth, der stets ein offenes Ohr für große wie kleine fachliche Herausforderungen hatte – und häufig auch direkt eine passende Lösung parat. Deine Ideen und Anregungen haben diese Arbeit maßgeblich beeinflusst, und an mancher Stelle hätte ich ohne deinen Rat sicher den Mut verloren.

Für die Bereitstellung und erste Aufbereitung der Messdaten, sowie für viele hilfreiche Gespräche und den kollegialen Austausch danke ich der gesamten Arbeitsgruppe des IBI – besonders Marija möchte ich hier hervorheben.

Mein besonderer Dank gilt auch Christine, Lasse, Nathalie, Nicolai und Steffi für das sorgfältige Korrekturlesen meiner Arbeit. Jule und Nicolai danke ich zudem ganz besonders für ihre spontane und tatkräftige Unterstützung in letzter Minute – sowohl vor der Abgabe als auch im Vorfeld der Disputation.

Abseits der fachlichen Unterstützung bin ich sehr dankbar für all die gemeinsamen Mittagessen (damals...), die unzähligen Kaffeepausen und die Gespräche auf dem Flur, die meinen Alltag in den letzten Jahren so angenehm und lebendig gemacht haben. Ihr, liebe Kolleginnen und Kollegen, habt diese Zeit wesentlich mitgeprägt.

Zu guter Letzt danke ich meiner Familie von Herzen. Meinen Eltern, die immer an mich glauben und mir den Rücken stärken, wann immer es nötig ist. Anni – danke, dass du die beste Schwester bist, die man sich wünschen kann. Hannes – du bist mein Fels in der Brandung. Danke für deine Geduld, deinen Beistand und deine Liebe. Und meine liebe Lotta – du zeigst mir jeden Tag aufs Neue, was wirklich zählt im Leben. Danke, dass du da bist.

Abstract

This thesis focuses on reconstruction of dynamic magnetic particle images. Magnetic particle imaging (MPI) is a tracer based imaging method with applications, e.g., in cardiovascular imaging, stroke detection and instrument tracking during interventions. It is characterized by a high spatial and temporal resolution, more particularly 3D measurements with a spatial resolution of 1 mm or less can be acquired with a temporal resolution of 46 frames per second. Thus, image reconstruction poses a highly dynamic problem. Mathematically, the MPI reconstruction problem is severely ill-posed already in a static setting and yields even more challenges for dynamic problems.

In this work, we propose the usage of a joint image reconstruction and motion estimation approach in order to improve reconstructed images by the additional motion information and improve computed displacement fields by enhanced image sequences. We use a variational problem formulation integrating a motion model that links image sequences and motion estimates. The problem is solved via an alternating minimization scheme, i.e., an image reconstruction subproblem and a motion estimation subproblem are solved alternately.

This thesis comprises a theoretical and a numerical part. From the theoretical perspective, we consider the forward operator of MPI and show a regularity property. Moreover, we show existence of a solution to the joint problem given this regularity assumption. Our theoretical results hold for a model incorporating the PDE describing the motion model as a hard constraint. However, numerically we consider a model incorporating the motion model as an additional penalty term. We observe that solutions of the unconstrained problems converge to a solution of the constrained problem.

From the numerical perspective, we derive and implement primal-dual algorithms to solve both the image reconstruction and the motion estimation subproblem. We perform extensive tests on simulated as well as measured data showing the applicability of the proposed approach. We observe superior performance compared to standard static methods that are not tailored to the dynamic reconstruction problem. Furthermore, the joint approach allows for subframe reconstruction, i.e., reconstruction of image sequences with a higher temporal resolution than the repetition time of the MPI scanner.

Kurzfassung

Diese Arbeit befasst sich mit der Rekonstruktion dynamischer Bildsequenzen aus der Magnetpartikelbildgebung. Die Magnetpartikelbildgebung (engl. magnetic particle imaging, MPI) ist ein tracer-basiertes Verfahren und wird zum Beispiel für kardiovaskuläre Bildgebung, das Aufspüren eines Schlaganfalls und Instrumententracking während Interventionen benutzt. MPI zeichnet sich durch eine hohe räumliche und zeitliche Auflösung aus. Insbesondere können 3D Messungen mit einer räumlichen Auflösung von weniger als 1 mm mit einer zeitlichen Auflösung von 46 Bildern pro Sekunde aufgenommen werden. Damit beschreibt die Bildrekonstruktion ein stark dynamisches Problem. Aus mathematischer Sicht ist bereits die Rekonstruktion statischer MPI Bilder ein stark schlecht gestelltes Problem, das durch die zusätzliche Zeitabhängigkeit im dynamischen Fall eine noch größere Herausforderung darstellt.
Hier wird ein gemeinsamer Bildrekonstruktions- und Bewegungsschätzungsansatz (engl. joint approach) betrachtet, um so die Bildrekonstruktion mithilfe zusätzlicher Bewegungsinformationen zu verbessern und die Qualität der berechneten Bewegungsfelder mittels verbesserter Bildsequenzen zu erhöhen. Dazu wird eine variationelle Problemformulierung genutzt, die ein Bewegungsmodell beinhaltet, um die Bildsequenzen und Bewegungsfelder zu verbinden. Das Problem wird mithilfe eines alternierenden Minimierungsalgorithmus gelöst, wobei abwechselnd das Bildrekonstruktionsproblem und das Bewegungsschätzungsproblem gelöst werden.
Diese Arbeit besteht aus einem theoretischen und einem numerischen Anteil. Auf der theoretischen Seite wird der MPI Vorwärtsoperator betrachtet und eine Regularitätsbedingung gezeigt. Damit ist es im Folgenden möglich, die Existenz einer Lösung des gemeinsamen Bildrekonstruktions- und Bewegungsproblems unter Annahme dieser Regularitätsbedingung zu zeigen. Außerdem wird bewiesen, dass Lösungen der unrestringierten Probleme, bei denen das Bewegungsmodell als zusätzlicher Strafterm eingefügt wird gegen eine Lösung des restringierten Problems konvergieren. Bei diesem ist das Bewegungsmodell als Nebenbedingung fast überall gefordert.
Im numerischen Teil der Arbeit werden primal-duale Algorithmen entwickelt und implementiert, um beide Teilprobleme zu lösen. Es folgen ausgiebige Tests sowohl auf simulierten Daten als auch auf Messdaten, um die Anwendbarkeit des betrachteten Ansatzes zu validieren. Im Vergleich zu statischen Algorithmen bringt der vorgeschlagene Ansatz bessere Ergebnisse hervor und erhöht sowohl die Qualität der rekonstruierten Bilder als auch die Genauigkeit der Bewegungsfelder. Mit dem betrachteten Ansatz ist es außerdem möglich, Subframe Rekonstruktionen zu berechnen, die sich durch eine höhere zeitliche Auflösung als die eigentlichen MPI Messungen auszeichnen.

Contents

List of acronyms

AAE	Averaged Angular Error
ADC	Analog-to-Digital Converter
ADMM	Alternating Direction Method of Multipliers
CLG	Combined Local-Global method
CT	Computed Tomography
DF	Drive-Field
FFP	Field-Free Point
FL	Fused Lasso, i.e., total variation plus L^1-norm
FOV	Field-Of-View
L1MC	Joint reconstruction algorithm using an L^1-data discrepancy term and the mass conservation motion model
L2MC	Joint reconstruction algorithm using an L^2-data discrepancy term and the mass conservation motion model
L1OF	Joint reconstruction algorithm using an L^1-data discrepancy term and the optical flow motion model
L2OF	Joint reconstruction algorithm using an L^2-data discrepancy term and the optical flow motion model
MC	Mass Conservation
MPI	Magnetic Particle Imaging
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
OF	Optical Flow
PDE	Partial Differential Equation
PDHG	Primal-Dual Hybrid Gradient method
PDHG-MC	PDHG algorithm for MC constrained motion estimation

PDHG-OF PDHG algorithm for OF constrained motion estimation

PET Positron Emission Tomography

PSNR Peak-Signal-to-Noise-Ratio

RESESOP Regularized Sequential Subspace Optimization algorithm

SNR Signal-to-Noise Ratio

SPDHG Stochastic Primal-Dual Hybrid Gradient method

SPECT Single Photon Emission Computed Tomography

SPION Super-Paramagnetic Iron-Oxide Nanoparticles

SSIM Structural Similarity Index Measure

TV Total Variation

1. Introduction

Nowadays, various medical imaging methods support diagnosis and treatment of illnesses and diseases. They allow for non-invasive imaging of the interior of the body, revealing hidden structures and pathological tissue. Having access to highly resolved multidimensional images of complex organs has brought a better understanding of critical diseases and revolutionized healthcare.

Some of the methods and applications thereof are well known to the public, as for example imaging bone fractures via X-ray, ligament injuries via Magnetic Resonance Imaging (MRI) or brain structures via Computed Tomography (CT). Among the less well known techniques we could name for example Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT). Whereas the first three mentioned approaches measure a property that is directly linked to the tissue under examination (*native imaging*), the last two named schemes apply a tracer to the human body and image its spatial distribution (*tracer-based imaging*). Tracer-based imaging is put to use especially in functional imaging and in cancer detection. These images do not provide any morphological insights. Here, only the tracer agent and not the tissue itself is imaged. Thus, they are often combined with native imaging methods in order to enhance interpretability of the resulting images.

Each technique is applied for specific tasks and they all have their pros and cons. For example in MRI, the patient experiences tissue heating as well as long scan times, during which he is exposed to an extremely noisy environment and might suffer from claustrophobia. In case of CT, there is exposure to ionizing radiation by X-rays and in case of PET and SPECT, the patient is exposed to gamma rays. There are several characteristics determining when and with which purpose a certain method is applied. Among these are the spatial and temporal resolution, the sensitivity and quantifiability, the cost and the amount of radiation applied.

In this thesis, we consider a rather new medical imaging technique called Magnetic Particle Imaging (MPI). In 2001, it was invented by Bernhard Gleich in the Philips Research Laboratories in Hamburg. MPI is a tomographic approach that can quantitatively map magnetic nanoparticle distributions in-vivo. It is characterized by high sensitivity, high spatial resolution and high imaging speed. In contrast to several clinically used schemes, MPI does not use any ionizing radiation. It is thus not harmful to the human body even under long-term considerations.

The research interest on MPI grew fast after the method was made public in 2005 [66]. First, the foci lay on optimizing the magnetic nanoparticles which are used as tracer material and investigating the particles' physics. Moreover, the scanner itself was up-scaled from a proof-of-concept device to a preclinical scanner with improved hardware. For a detailed historical perspective on achieved milestones, we refer to [91]. For an extensive review on recent developments and major theoretical results, we refer to the surveys [87, 135]. MPI is tailored for applications requiring fast dynamic imaging such as, e.g., blood flow visualization in the case of coronary artery diseases. Other future applications include cancer detection [143], detection of stroke and other neurological pathologies like hemorrhage, tumors and inflammatory processes [108, 109], stem cell monitoring [137, 147], localization of medical instruments during vascular interventions [70] or any application where tracers are used. To date, MPI is still in the preclinical phase.

Let us now consider image reconstruction in medical imaging applications. Image reconstruction refers to the task of recovering the full dimensionality of an object from the raw measured signal. It thus plays an important role for obtaining interpretable image data, which can then be handed to the physicians. In a first step, we have to find and understand the forward model A of the considered imaging device. This model summarizes the physical principles behind an imaging method. Measured data u (also called our observation) occur when this forward model is applied to some cause, e.g., human tissue denoted by c. We have to understand and describe the generation of data u from known forward model and cause. This task is referred to as the *direct problem.* Then, in a second step, we consider the reconstruction of images, i.e., the reconstruction of the cause c from known forward model A and data u by

$$Ac = u.$$

Recovering c poses an *inverse problem.* Inverse problems in imaging are often ill-posed, meaning that either a solution does not exist, is not unique or, and this is the most problematic case, has a solution that does not depend continuously on the data. In this case, small changes of the data lead to huge differences of the solution. A wealth of theoretical results and numerical methods are available for inverse problems in a static setting and can be found in textbooks, see, e.g., [55].

Often, cause and observation show a time-dependent behavior such that dynamic inverse problems occur naturally in medical imaging applications. Imagine for example a CT scan of the torso, where the source of x-rays is rotated around the patient inside in order to obtain data from different directions. Thus, the observation shows a time-dependent behavior. Moreover, the patient inside the scanner breathes such that also the cause is time-dependent.

The research field of dynamic inverse problems has been strongly driven by tomographic modalities in the past. Applying static reconstruction algorithms often yields severe motion artifacts caused by the dynamic nature of the object. Including time simply as another dimension changes the characteristics of the respective inverse problem. Thus, we need tailored reconstruction approaches taking into account the dynamic nature of observation and cause. However, while there is extensive generalized theory on static inverse problems, no general regularization framework for dynamic inverse problems has been established so far. Various different approaches are proposed, based on, e.g., temporal smoothness assumptions for a fairly general setting not requiring any additional information about the motion itself [126, 127]. Other approaches include dynamic programming techniques [85], explicit deformation models [71] or operator inexactness [22].

An extensive survey on variational approaches based on parametrised temporal models can be found in [74]. In [81], a collection of works in dynamic inverse problems in imaging and parameter identification was published providing an overview on recent developments. Works focusing on modelling, analysis and regularization of dynamic inverse problems are collected in the special issue [129].

Depending on the characteristic features and challenges of specific imaging devices, different algorithmic approaches have been proposed having in mind particularly those situations. We name a few such approaches, although the list is not intended to be exhaustive. Exact analytic methods have been proposed, e.g., for compensation of respiratory motion during CT scans [45] and for compensation of more general deformations during CT [49, 121]. For dynamic CT, approximate inversion formulas are also obtainable [6, 72, 82, 83]. Another option are iterative methods [23, 78, 80]. Further, machine-learning based algorithms applied to, e.g., artifact reduction in MRI [73, 96] and joint image reconstruction and motion estimation on PET data [100] have also been proposed. Furthermore, variational approaches are also popular, they can be divided into the class exploiting a motion model, the ones exploiting a deformable template and others, which rely on none of those. In [104], an alternating variational approach is used to solve a model incorporating the so-called 5D respiratory motion model in order to reconstruct 4D lung CT image sequences. When acquiring motion parameters prior to or within the image reconstruction scheme, motion models can be incorporated explicitly [34, 35, 72, 74, 105, 107, 117]. The authors of [43] propose a joint model for image reconstruction and motion estimation applicable to spatiotemporal imaging by sequential indirect image registration. Other approaches based on deformable templates are proposed in [42, 99].

In this thesis, we consider a variational method incorporating a motion model explicitly. More particularly, we consider a joint image reconstruction and motion estimation approach, i.e., an image sequence and a displacement field, i.e., the motion in between the images, are obtained simultaneously. The idea behind this scheme is that both tasks endorse each other, i.e., incorporating motion estimates improves the resulting image quality, and incorporating better images then again improves the quality of the motion estimates. Instead of formulating one optimization problem for the image reconstruction task and then another one for motion estimation, we formulate one single optimization problem incorporating prior information from the imaging device as well as a motion model. The optimization problem is then solved for both unknowns, i.e., image sequence and motion. As we will see, the optimization problem has a complex structure. It is non-differentiable and convex only with respect to one variable, but not biconvex with respect to both. We make use of concepts from convex optimization in order to solve the problem. Tomasi and Kanade introduced the idea of joint motion estimation and image reconstruction in [138]. A joint model for gated cardiac CT solved by the conjugate gradient method was proposed in [65]. A closely related topic is joint motion estimation and image deblurring, a topic which is covered by a variational approach in [12]. The approach used in this thesis was proposed in [35, 50] for two dimensional image sequences and applied to temporal inpainting and cell tracking.

Coming back to the special case of MPI reconstruction, we note that the image reconstruction task is already severely ill-posed in the static case [56, 90]. It becomes even more ill-posed in the dynamic case. However, as indicated by the potential applications listed above, reconstructing dynamic image sequences is of high interest. Few works on dynamic MPI reconstruction exist so far. From a modeling point of view, the setting was analyzed in [28]. A reconstruction approach designed for periodic motion using data binning into virtual frames was considered in the single and multi-patch setting in [63, 64]. This approach is limited to very specific classes of motion, whereas in [29] temporal splines are applied in order to compensate non-periodic motion and allow for a high temporal resolution. In [114], the authors suggest to take into account dynamics as model inexactness and apply the Regularized Sequential Subspace Optimization algorithm (RESESOP) to solve the problem numerically. Motion estimation algorithms have been proposed for extracting flow information in a second step from previously reconstructed image sequences in [60].

As mentioned before, we aim at obtaining image sequences and motion estimates simultaneously in this thesis. A first step in this direction was taken in [88], which can be seen as a proof of concept for this approach applied to MPI reconstruction. The authors apply a joint scheme to two dimensional MPI image sequences. In contrast to this motivational work, we particularly consider the three dimensional case and focus on the theoretical justification of applying the approach to MPI.

1.1. Outline of the thesis

We start with a recapitulation of some concepts and results from functional analysis, convex optimization and inverse problems, which are relevant for this thesis in Chapter 2. Thereby, we pay special attention to function spaces and function space embeddings as well as the concept of duality, in particular Fenchel-Rockafellar duality.

Afterwards, we consider MPI in detail in Chapter 3. We introduce this novel imaging method and derive the static forward model in Section 3.1. Thereby, we observe compactness of the forward operator, showing the ill-posedness of the image reconstruction task. Our main achievement in this section is the proof of Theorem 3.1, which states that the MPI forward model is non-vanishing. This property is crucial for the main proof of Theorem 5.1. Furthermore, we extend the forward model from static to dynamic particle distributions and consider different timescales in MPI in Section 3.2. Finally, we conclude this chapter by introducing different approaches to static image reconstruction and compare them on synthetic data in Section 3.3.

Chapter 4 deals with the motion estimation task. We give a brief introduction into the general topic and obtain two different motion models, namely the Optical Flow and Mass Conservation motion model in Section 4.1. Afterwards, we introduce the main classes of motion estimation algorithms and derive primal-dual algorithms for both aforementioned motion models in Section 4.2. Further, we address the well-known challenge of large displacements by deriving a multiscale approach, partially combined with image warping. The algorithmic solutions are tested extensively on academic test cases showing a solid performance at reasonable costs in Section 4.3.

Chapter 5 is dedicated to the joint image reconstruction and motion estimation task applied to three dimensional MPI. In Section 5.1, we show the well-definedness of the

joint problem formulation by proving existence of a minimizer in Theorem 5.1. Based on [35, 50], we extend the proof to three dimensional image sequences, allow for a time-dependent forward operator and consider more general regularization terms. Theorem 5.1 shows existence of a solution to a problem constrained by the motion model. Moreover, we show convergence of solutions of the unconstrained minimization problem, where the constraint is included in the minimization problem as an additional regularization term. In Section 5.2, we briefly analyze how our approach defines a regularization method to a nonlinear problem. Section 5.3 considers numerical treatment of the joint optimization problem by an alternating minimization approach.
In Chapter 6, we present numerical examples of the joint approach discussed in the previous chapter. In Section 6.1, we consider experiments on simulated data, showing the applicability to MPI data and the advantages compared to a two-step reconstruction scheme, i.e., reconstruction of image sequences first followed by the computation of motion. In Section 6.2 and Section 6.3 we treat real measurements of an in-vitro rotation phantom and in-vivo data of the cardiovascular system of a mouse, respectively, underlining the applicability in practice. Section 6.4 is dedicated to the topic of subframe reconstruction in order to further improve the temporal resolution of MPI images.
We conclude this thesis and discuss the results in Chapter 7. Moreover, we provide an outlook to possible future research directions.

1.2. Contribution

We now summarize the main contributions of this thesis. A more detailed presentation of the contributions can be found in the introductions of the corresponding chapters.
In this work, we develop a method for joint motion estimation and image reconstruction for a general 3D plus time setting for time-dependent linear forward operators and fairly general regularization terms in Chapter 5. We thus extend the work of [35, 50] to this more general setting. We show well-definedness of the joint problem formulation and prove convergence of solutions of the unconstrained optimization problems to a solution of the constrained problem. The theoretical aspect of this work is complemented by numerical experiments in Chapter 6. We perform extensive studies on simulated as well as measured data sets analyzing the enhancement of the image quality and the motion estimate accuracy by the proposed joint algorithm. The implementation of the corresponding algorithms in Matlab was performed while working on this thesis. Preceding the joint approach, we consider MPI reconstruction and the motion estimation task separately. Doing this, we perform analytical investigations of the static MPI forward operator ensuring a necessary regularity property in Chapter 3. More particularly, we show that the forward operator is non-vanishing over a compact time interval, if this time interval includes a non-stationary point of the field-free point trajectory. This enables us to apply the developed theory on joint image reconstruction and motion estimation to a large class of MPI scanners. In order to obtain a motion estimation algorithm applicable to the joint setting, we formalize multiscale primal-dual algorithms and include a warping procedure in Chapter 4. Moreover, we propose different approaches to improve the robustness to noise of those algorithms to make them applicable to potentially noisy MPI data.

2. Preliminaries

This chapter summarizes mathematical concepts used throughout this thesis. Although the topics of the different sections are highly connected, we split the chapter into several sections for readability. We start with some fundamental definitions about function spaces and theory about embeddings of function spaces in Section 2.1. The section also introduces the weak and weak-* topologies as additional topologies. The most essential statements for the analysis in later chapters are the Lemma of Aubin-Lions and the Theorem of Banach-Alaoglu. In Section 2.2, we give an overview on tools and ideas from convex optimization. We define convexity, introduce subdifferentiability in order to weaken the assumption of differentiable functionals and derive the setting of Fenchel-Rockafellar duality. Proximal operators are introduced and computed for specific functions that are used in following chapters. Combining results of the first two sections will enable us to prove existence of a minimizer for the joint approach in Chapter 5. Section 2.3 completes the chapter. In this section, we briefly introduce inverse problems and their challenges and define regularization methods for such problems.

2.1. Functional analysis

The main theoretical result of Chapter 5 is given by the existence of solutions for the joint problem formulation. In order to prove the corresponding theorem, we need some definitions, embedding theorems and other fundamental results from functional analysis that are stated in the following. This section is mostly based on [3, 31], to which we also refer for a more comprehensive introduction to the main topics. Definitions are taken from [3, 31] if not mentioned explicitly.
We start by introducing the very basic concept of the space of linear operators and the dual space.

Definition 2.1: Linear operators and the dual space

Let X be a normed space. The **space of linear operators** mapping from X to a space Y is denoted by $\mathcal{L}(X,Y)$, i.e.,

$$\mathcal{L}(X,Y) = \{A : X \to Y \,|A \text{ linear and continuous}\}.$$

The space $\mathcal{L}(X,\mathbb{R})$ of all linear functionals mapping to $\mathbb{R}$ is called the dual space of X and denoted by X^*.

Considering two spaces, we will often be interested in their relation in terms of embeddings as defined in the following, see [30].

Definition 2.2: Embeddings

Let X and Y be Banach spaces with associated norms $\|\cdot\|_X$ and $\|\cdot\|_Y$ and let $X \subset Y$. We call X

- **continuously embedded** in Y if the inclusion map $X \to Y, u \mapsto u$ is continuous, i.e.,

  $$\|u\|_Y \leq c \|u\|_X$$

  for all $u \in X$ and a constant $c > 0$. We denote this by $X \hookrightarrow Y$.

- **compactly embedded** in Y if the inclusion is a compact operator, i.e., any bounded sequence has a subsequence that is Cauchy in Y. We denote this by $X \Subset Y$.

We now derive first basic and then more complex function spaces and state some properties, including main embedding theorems and the dual spaces, for each of them. In the following, $\Omega \subset \mathbb{R}^n$ denotes an open and bounded domain. Additional assumptions are mentioned explicitly.

Definition 2.3: Lebesgue spaces

Let F be a measurable function and $1 \leq p < \infty$. The space

$$L^p(\Omega) = \left\{ F : \Omega \to \mathbb{R} \,\middle|\, \int_\Omega |F(u)|^p \,\mathrm{d}u < \infty \right\}$$

is called **Lebesgue space** and equipped with the norm

$$\|F\|_{L^p(\Omega)} = \left(\int_\Omega |F(u)|^p \,\mathrm{d}u \right)^{\frac{1}{p}}.$$

Note that, in order to obtain a norm and not only a semi-norm, we identify all functions with each other that coincide almost everywhere. That is, we define the space with the equivalence relation

$$F = G \;\text{ in }\; L^p(\Omega) :\Leftrightarrow F = G \;\text{ almost everywhere.}$$

The Lebesgue space for $p = \infty$ contains essentially bounded functions, i.e.,

$$L^\infty(\Omega) = \left\{ F : \Omega \to \mathbb{R} \,\middle|\, \sup_{u \in \Omega \setminus \mathcal{N}} |F(u)| < \infty \right\},$$

for $\mathcal{N}$ a set of zero measure.

Lebesgue spaces are Banach spaces, see the Fischer-Riesz theorem with proof for example in [31, Theorem 4.8]. The dual space of $L^p(\Omega)$ for $1 < p < \infty$ is given by $L^q(\Omega)$ with

$$\frac{1}{p} + \frac{1}{q} = 1,$$

and L^p spaces are reflexive for $1 < p < \infty$ [31, Theorem 4.10]. The dual space to $L^1(\Omega)$ is given by $L^\infty(\Omega)$, whereas the dual space to $L^\infty(\Omega)$ contains $L^1(\Omega)$ but is strictly bigger [31, pp. 99-102]. When considering norm estimates in Lebesgue spaces, there are some important inequalities that have to be mentioned and are thus stated in the following for completeness.

Lemma 2.1: Hölder inequality

Let $1 \leq p, q \leq \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$ and $f \in L^p(\Omega)$, $g \in L^q(\Omega)$. For the product fg it holds that $fg \in L^1(\Omega)$ and

$$\|fg\|_{L^1(\Omega)} \leq \|f\|_{L^p(\Omega)} \|g\|_{L^q(\Omega)} .$$

More generally, for $m \in \mathbb{N}$ and $f_i \in L^{p_i}(\Omega)$ for $i = 1, ..., m$ with $p_i \in [1, \infty]$ and $q \in [1, \infty]$ such that

$$\sum_{i=1}^{m} \frac{1}{p_i} = \frac{1}{q},$$

it holds that $\prod_{i=1}^m f_i \in L^q(\Omega)$ and

$$\left\| \prod_{i=1}^{m} f_i \right\|_{L^q(\Omega)} \leq \prod_{i=1}^{m} \|f_i\|_{L^{p_i}(\Omega)} .$$

For a proof see for example [3, Lemma 3.18].

Lemma 2.2: Minkowksi inequality

Let $1 \leq p \leq \infty$ and $f, g \in L^p(\Omega)$. Then $f + g \in L^p(\Omega)$ and

$$\|f + g\|_{L^p(\Omega)} \leq \|f\|_{L^p(\Omega)} + \|g\|_{L^p(\Omega)} .$$

A proof can be found in [3, Lemma 3.20].

Lemma 2.3: Lebesgue embedding

Let $\Omega \subset \mathbb{R}^n$ be bounded and $1 \leq p < q \leq \infty$. Then

$$L^q(\Omega) \hookrightarrow L^p(\Omega).$$

Proof. Let $F \in L^q(\Omega)$. Then

$$\|F\|_{L^p(\Omega)} = \|\chi_\Omega F\|_{L^p(\Omega)} \leq \|\chi_\Omega\|_{L^r(\Omega)} \|F\|_{L^q(\Omega)} = |\Omega| \, \|F\|_{L^q(\Omega)} < \infty$$

for $\frac{1}{p} = \frac{1}{r} + \frac{1}{q}$ by using the Hölder inequality. The function χ_Ω denotes the characteristic function of the set Ω. $\square$

Definition 2.4: Sobolev spaces

Let $F : \Omega \to \mathbb{R}$ be measurable and $k \in \mathbb{N}$, $1 \leq p \leq \infty$. The space

$$W^{k,p}(\Omega) = \{F \in L^p(\Omega) \,|\, \partial^\alpha F \in L^p(\Omega) \;\; \forall |\alpha| \leq k\}$$

is called **Sobolev space** and contains functions whose weak derivatives up to order k are again in $L^p(\Omega)$. The space is a Banach space equipped with the norm

$$\|F\|_{W^{k,p}(\Omega)} = \left(\sum_{|\alpha| \leq k} \|\partial^\alpha F\|^p_{L^p(\Omega)} \right)^{\frac{1}{p}}.$$

The space $W^{k,p}(\Omega)$ is reflexive for $1 < p < \infty$ and Ω bounded [31, Proposition 8.1.]. By $W_0^{1,p}(\Omega)$ we denote the closure of $\mathcal{C}_0^1(\Omega)$ in $W^{1,p}(\Omega)$. This space is reflexive for $1 < p < \infty$. The dual of $W_0^{1,p}(\Omega)$ is denoted by $W^{-1,q}(\Omega)$.

By $H^1(\Omega)$ we denote the (Hilbert) space $W^{1,2}(\Omega)$. The following dimension dependent embedding theorem is crucial for the proof in Chapter 5.

Theorem 2.1: Sobolev embedding theorem

Let $\Omega \subset \mathbb{R}^n$ with Lipschitz boundary, $m \geq 1$ and $1 \leq p < \infty$. The following continuous embeddings hold:

- If $mp \geq n$, then

$$W^{m,p}(\Omega) \hookrightarrow L^q(\Omega) \qquad \text{for } \; p \leq q \leq \infty.$$

- If $mp < n$, then

$$W^{m,p}(\Omega) \hookrightarrow L^q(\Omega) \qquad \text{for } \; p \leq q \leq \frac{np}{n - mp}.$$

Different conditions on Ω as well as more details, references and the proof can be found in [1, Theorem 4.12].

We now introduce a function space popular in imaging for two reasons: first, the space has desirable properties (and therefore appears in many results in the following chapters) and second, the space contains image functions naturally. The following definition is based on the representation in [3, Theorem 6.26].

Definition 2.5: Functions of bounded variation

Let $F : \Omega \to \mathbb{R}$ be measurable. The **space of functions with bounded variation** $BV(\Omega)$ is defined by

$$BV(\Omega) = \left\{ F \in L^1(\Omega) \,\middle|\, |F|_{BV(\Omega)} < \infty \right\}$$

for

$$|F|_{BV(\Omega)} = \sup_{\phi \in \mathcal{C}_0^\infty(\Omega,\mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \leq 1} \int_\Omega F \nabla \cdot \phi \mathrm{d}x$$

where $|F|_{BV(\Omega)}$ is called the **Total Variation (TV)** of F. The TV does not denote a norm but only a semi-norm, such that the space $BV(\Omega)$ is equipped with the following norm in order to obtain a Banach space

$$\|F\|_{BV(\Omega)} = |F|_{BV(\Omega)} + \|F\|_{L^1(\Omega)}$$

Remark. For a function $F \in W^{1,1}(\Omega)$, the total variation reduces to

$$|F|_{BV(\Omega)} = \int_\Omega |\nabla F| \, \mathrm{d}x.$$

Another property, that will be of interest for us, is the lower semicontinuity of the total variation. A proof, as well as more interesting properties can be found for example in [5, Remark 3.5].

The following characterization is needed to apply the theorem of Banach-Alaoglu.

Lemma 2.4: BV as a dual space

The space $BV(\Omega)$ is the dual space of a separable Banach space Y.

The construction of the space and more details can be found in [5, Remark 3.12]. While we now know that $BV(\Omega)$ defines a dual space, very little can be said on the dual space of $BV(\Omega)$. We now consider how the space of bounded variation embeds into Lebesgue spaces.

Theorem 2.2: Embedding of BV

Let $\Omega \subset \mathbb{R}^n$ with Lipschitz boundary. Then,

$$BV(\Omega) \hookrightarrow L^l(\Omega) \qquad \text{for } 1 \leq l \leq \frac{n}{n-1},$$
$$BV(\Omega) \Subset L^l(\Omega) \qquad \text{for } 1 \leq l < \frac{n}{n-1}.$$

A proof can be found in [5, Theorem 3.49]. As we consider time and space dependent functions in this work, we introduce another family of function spaces, see [131] for more details on the spaces.

Definition 2.6: Bochner spaces

Let Y be a Banach space with norm $\|\cdot\|_Y$, $1 \leq p \leq \infty$ and F continuous. Then

$$L^p(0,T;Y) = \left\{ F : [0,T] \to Y \,\middle|\, \|F\|_{L^p(0,T;Y)} < \infty \right\}$$

is called a **Bochner space**. The associated norm is defined by

$$\|F\|_{L^p(0,T;Y)} = \left(\int_0^T \|F(t)\|_Y^p \, \mathrm{d}t \right)^{\frac{1}{p}} \quad \left(= \operatorname{ess} \sup_{0<t<T} \|F(t)\|_Y \ \text{ if } \ p = \infty \right).$$

The dual space of $L^p(0,T;Y)$ for $1 < p < \infty$ is given by $L^q(0,T;Y^*)$ with Y^* being the dual of Y and

$$\frac{1}{p} + \frac{1}{q} = 1.$$

The dual space to $L^1(0,T;Y)$ is given by $L^\infty(0,T;Y^*)$. A proof and more details on duality for Bochner spaces can be found in [39]. For a reflexive Banach space Y, the Bochner space $L^p(0,T;Y)$ is reflexive for $1 < p < \infty$.

Lemma 2.5: Bochner space embeddings

- Let $1 \leq p \leq \infty$ and $X \hookrightarrow Y$. Then

$$L^p(0,T;X) \hookrightarrow L^p(0,T;Y).$$

- Let $1 \leq p < q \leq \infty$. Then

$$L^q(0,T;X) \hookrightarrow L^p(0,T;X).$$

For a proof, see [50, Lemma 2.1.23.].
Unlike for Lebesgue spaces, where the compact embedding $L^q(\Omega) \Subset L^p(\Omega)$ for $1 \leq p \leq q$ and Ω bounded exists naturally, there is no compact embedding $L^q(0,T;X) \Subset L^p(0,T;Y)$ in general for $X \Subset Y$. Instead, we utilize the relative compactness, i.e., the compactness of the closure of such embeddings as proven in the Lemma of Aubin-Lions.

Theorem 2.3: Lemma of Aubin-Lions

Let X, Y and Z be Banach spaces with $X \subseteq Y \subseteq Z$ and $X \Subset Y$ and $Y \hookrightarrow Z$. Let either $q = 1$ and $1 \leq p < \infty$ or $q > 1$ and $1 \leq p \leq \infty$. Moreover, let f_n be a sequence of bounded functions in $L^p(0,T;X)$ and let $\frac{\partial}{\partial t} f_n$ be bounded in $L^q(0,T;Z)$.
Then f_n is relatively compact in $L^p(0,T;Y)$.

The original proof by Aubin [11] assumes reflexive Banach spaces, the version without this additional assumption was proved by Simon in [131]. At this point, we have introduced all function spaces needed throughout this work and stated their main properties. We now turn our attention to different topologies on such spaces. The standard convergence property with respect to the norm of a Banach space can be too restrictive in many cases. The main difficulty is to choose a convergent subsequence from a minimizing sequence, as unit balls are in general not pre-compact with respect to the norm in infinite dimensional spaces. Therefore, other topologies are considered.
There are different topologies on dual spaces Y^*. First, there is the usual strong topology associated to the norm of Y^*. In addition, we have the weak topology.

Definition 2.7: Weak convergence

Let Y be a Banach space with dual space Y^*. A sequence $u_n \in Y$ **converges weakly** to $u \in Y$ if

$$\langle u_n, v\rangle \to \langle u, v\rangle \qquad \text{for } n \to \infty$$

for all v in Y^*, where $\langle \cdot, \cdot\rangle$ denotes a dual product here. We denote weak convergence by $u_n \rightharpoonup u$.

For reflexive Banach spaces, it can be shown that the unit ball is weakly compact (e.g., [3, Theorem 6.10]). For non-reflexive spaces however, this does not hold.
The third topology we define is the weak-* topology. This topology has more compact sets compared to the previously mentioned ones, which makes it important to consider.

Definition 2.8: Weak-* convergence

Let Y be a Banach space with dual space Y^*. We say a sequence $v_n \in Y^*$ **weak-*-converges** to $v \in Y^*$ if

$$\langle u, v_n\rangle \to \langle u, v\rangle \qquad \text{for } n \to \infty$$

for all u in Y. Again, $\langle \cdot, \cdot\rangle$ denotes a dual product in this setting. We denote weak-* convergence by $v_n \overset{*}{\rightharpoonup} v$.

The most essential property regarding the weak-* topology is stated in the following theorem.

Theorem 2.4: Banach-Alaoglu

Let Y be a Banach space and Y^* its dual space. The closed ball B_ε

$$B_\varepsilon = \{v \in Y^* : \|v\|_{Y^*} \leq \varepsilon\}$$

is compact with respect to the weak-* topology.

A proof for $\varepsilon = 1$ can be found in [31, Theorem 3.16].

2.2. Convex optimization

This section gives an overview of the main tools and ideas of convex optimization. These will be used in succeeding chapters to prove existence and uniqueness of minimizers, but also to derive algorithms to compute those. The results mentioned in this section, as well as further material can be found in standard references such as [19, 25, 54, 120], the seminal work of Rockafellar [119] or the more recently published textbooks [17, 30, 44]. We present extensive examples in this section, computing many specific functionals needed in succeeding applied chapters. Thereby, we can simply refer to those examples later on. We start by giving some basic definitions that will be needed throughout this thesis. In the following, we denote $\mathbb{R}_\infty := \mathbb{R} \cup \{+\infty\}$.

Definition 2.9: Convexity

A functional $F : X \to \mathbb{R}_\infty$ on a normed space X is called **convex** if for all u, $v \in X$ and $\lambda \in [0,1]$ it holds that

$$F(\lambda u + (1-\lambda) v) \leq \lambda F(u) + (1-\lambda) F(v).$$

It is called **strictly convex** if that inequality holds strictly for $\lambda \in (0,1)$ and $u \neq v$.

There are some simple examples for convex functionals, which play an important role in the context of inverse problems and regularization theory.

Example 2.1: Convex functionals

1. Norm functionals.
   The norm $\|\cdot\|_X$ on a normed space X is convex due to the absolute homogeneity and triangle inequality, i.e., for $u, v \in X$ we have

   $$\|\lambda u + (1-\lambda) v\|_X \leq \|\lambda u\|_X + \|(1-\lambda) v\|_X = \lambda \|u\|_X + (1-\lambda) \|v\|_X .$$

   For a simple illustration, cf. Figure 2.1a, which depicts the absolute value function.

2. Indicator functionals on convex sets.
   Let $K \subset X$ be a convex set. Then the corresponding indicator functional is defined by

   $$\mathbb{I}_K(u) = \begin{cases} 0, & u \in K \\ \infty, & \text{else} \end{cases} .$$

   As K is a convex set, $\lambda u + (1-\lambda) v \in K$ for $u, v \in K$ and thus the functional is convex. For illustration, cf. Figure 2.1b, which depicts the indicator function over a closed interval.

3. Composition with a linear map.
   Let $F : X \to \mathbb{R}_\infty$ be a convex functional and $A : \operatorname{dom}(A) \to X$ be a linear operator. The composition

   $$F \circ A : \operatorname{dom}(A) \to \mathbb{R}_\infty$$

   is then convex as

   $$\begin{aligned}(F \circ A)(\lambda u + (1-\lambda) v) &= F(\lambda A u + (1-\lambda) A v) \\ &\leq \lambda F(Au) + (1-\lambda) F(Av).\end{aligned}$$

4. Linear combinations of convex functionals.
   Let F and G be convex functionals on a normed space X and $\alpha, \beta \geq 0$. Then the functional $\alpha F + \beta G$ is also convex.

Convex functionals have several nice properties. This makes them useful in the context of optimization problems. We name the main results, which are helpful in the following chapters.

Theorem 2.5: Minima of convex functionals

Let $F : X \to \mathbb{R}_\infty$ be a convex functional on a reflexive Banach space X. Then each local minimum is a global one.
If F is even strictly convex, then F has at most one minimizer.

A proof can be found for example in [19, Proposition 3.1.1.].

Definition 2.10: Lower semicontinuity

A functional $F : X \to \mathbb{R}_\infty$ on a normed space X is called **(sequentially) lower semicontinuous** in $u \in X$ if for all sequences u_n with $\lim_{n\to\infty} u_n = u$, it holds that

$$F(u) \leq \liminf_{n\to\infty} F(u_n).$$

It is called **weakly lower semicontinuous** if that inequality holds for all sequences u_n converging weakly to u.

Definition 2.11: Coercivity

A functional $F : X \to \mathbb{R}_\infty$ on a Banach space X is called **coercive** if for sequences u_n with $\|u_n\|_X \to \infty$ for $n \to \infty$, it holds that

$$F(u_n) \to \infty \quad \text{for } n \to \infty.$$

The following Theorem 2.6 states what is also known as the direct method of calculus of variations and is typically used to show existence of minimizers for functionals. The idea is as follows. Consider the minimization problem

$$\min_{u\in X} F(u).$$

After establishing a bound from below, one constructs a minimizing sequence $u_n \in X$. We then show that this sequence lies in a sequentially compact set such that there is a convergent subsequence with respect to a certain topology. We thus have a candidate for a minimizer. Proving lower-semicontinuity with respect to the same topology shows that the candidate indeed is a minimizer. The choice of topology is crucial. The weaker the topology, the easier it is to obtain a sequentially compact set. By contrast, it is more difficult to obtain lower semicontinuity for weaker topologies. Choosing the weak topology on a space X enables the following line of argumentation. If F is coercive, we obtain a uniform bound on the minimizing sequence u_n. If moreover, X is reflexive, the sequence lies in a ball of finite radius which is weakly compact. Thus, there exists a subsequence weakly convergent to an element u_0. If F is weakly lower semicontinuous, then u_0 is a minimum point of F.

Theorem 2.6: Existence of solutions

Let X be a reflexive Banach space and $F : X \to \mathbb{R}_\infty$. If F is bounded from below, weakly lower semicontinuous and coercive, then there exists a solution u^* to the minimization problem

$$\min_{u \in X} F(u) .$$

A detailed proof can be found for example in [30, Theorem 6.17/ Section 6.2.1].

Remark. A similar result holds for dual spaces X^* of separable normed spaces under the assumption that F is weak-* lower semicontinuous [30]. The weak-* compactness is then realized by Banach-Alaoglu (Theorem 2.4).

Remark. If F is convex, F is lower semicontinuous if and only if F is weakly lower semicontinuous such that the assumptions reduce to F being convex, lower semicontinous and coercive. If $F : X \to \mathbb{R}_\infty$ is strictly convex, lower semicontinuous and coercive, this guarantees existence of a unique minimizer.

In the case of differentiable functionals, convexity can also be expressed as a condition to the derivative of the functional. Moreover, necessary and sufficient conditions for solutions to the corresponding minimization problems can be formulated in terms of the derivative of first and second order. However, we do not want to limit ourselves to differentiable functions in the following and directly introduce a more general idea of differentiability.

Definition 2.12: Subdifferential

Let $F : X \to \mathbb{R}_\infty$ be a convex functional on a normed space X. The **subgradient** at a point $u \in X$ is defined as an element $x^* \in X^*$ that fulfills

$$F(u) + \langle x^*, v - u \rangle \leq F(v) \quad \forall \ v \in X .$$

The **subdifferential** $\partial F(u)$ at the point $u \in X$ contains all subgradients at u.

Remark. If the functional F is Gâteaux-differentiable at a point u, then the subdifferential at this point contains only the Gâteaux-derivative of F at u.

Example 2.2: Subdifferentials of convex functionals

1. The absolute value function.
   Let $f(u) = |u|$. Then f is differentiable for $u \neq 0$. For $u = 0$, the inequality describing the subdifferential reads

   $$\langle x^*, y \rangle \leq |y| \quad \forall y \in \mathbb{R},$$

   which yields $\partial f(0) = \left\{ x^* \middle| x^* \in [-1, 1] \right\}$. The subdifferential is depicted in Figure 2.1a.

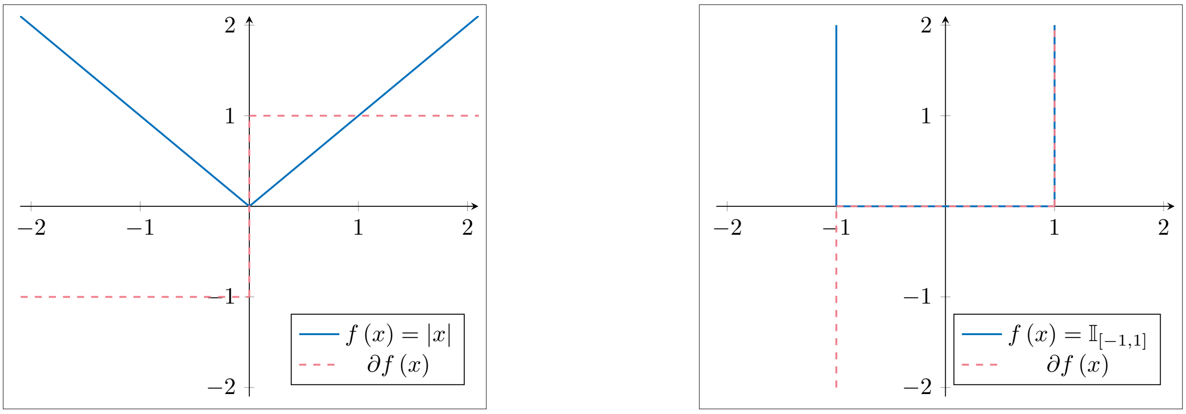
2. Indicator functional on the interval $[-1, 1]$.
   The indicator functional is defined by

$$\mathbb{I}_{[-1,1]}(u) = \begin{cases} 0, & u \in [-1, 1] \\ \infty, & \text{else} \end{cases}.$$

   The function is differentiable for $u \in (-1, 1)$ with $\partial(\mathbb{I}_{[-1,1]})(u) = \{0\}$. For $u = -1$, a subgradient x^* has to fulfill the inequality

$$0 + \langle x^*, y \rangle \leq \mathbb{I}_{[-1,1]}(y) = \begin{cases} 0, & y \in [-1, 1] \\ \infty, & \text{else} \end{cases},$$

   for all $y \in \mathbb{R}$. This leads to $-\infty < x^* \leq 0$.
   Similarly, $\partial(\mathbb{I}_{[-1,1]})(1) = \left\{x^* \middle| x^* \in (0, \infty)\right\}$. The subdifferential is illustrated in Figure 2.1b.



(a) The absolute value function.

(b) The indicator function.

Figure 2.1.: Examples for convex functions and their subdifferentials.

A necessary and sufficient condition for minimizers is defined in the following theorem.

Theorem 2.7: Fermats theorem

Let $F : X \to \mathbb{R}_\infty$ be a convex functional on a normed space X. Then $u^* \in X$ is a (global) minimizer of F if and only if $0 \in \partial F(u^*)$.

Proof. Let u^* be a minimizer of F. This holds if and only if, $F(u^*) \leq F(u)$ for all $u \in X$. This is equivalent to

$$F(u^*) + \langle 0, u - u^* \rangle \leq F(u) \quad \forall u \in X,$$

which describes the subdifferential inequality with 0 plugged in as a subgradient. □

2.2.1. Fenchel-Rockafellar duality

In some cases, a minimization problem might be difficult to handle but can be reformulated into an equivalent problem that is easier to solve. This section summarizes the main tools and ideas needed to derive the dual problem as well as saddle point formulations. Our examples mainly consist of functionals that are often used in imaging applications and will be needed in succeeding chapters. We refer to [30] for a rather applied introduction to the topic.

Definition 2.13: Fenchel conjugate

Let $F : X \to \mathbb{R}_\infty$ be a proper functional on a real Banach space X, $u \in X$ and $v \in X^*$. The **Fenchel conjugate** F^* is defined by

$$F^* : X^* \to \mathbb{R}_\infty, \quad F^*(v) = \sup_{u \in X} \{\langle v, u\rangle - F(u)\}.$$

Remark. For F proper and $u \in X$, $v \in X^*$ it holds that

$$\langle v, u\rangle \leq F(u) + F^*(v).$$

This inequality is called **Fenchel inequality**. With the help of this equality, one can state the fundamental relation between subdifferential and Fenchel conjugate by

$$v \in \partial F(u) \Leftrightarrow \langle v, u\rangle = F(u) + F^*(v).$$

This relation can be used to replace the subdifferential of a possibly complicated functional by the one of a simpler conjugate functional.

Example 2.3: Fenchel conjugates

1. Norm functionals.
   Let $F(u) = \varphi(\|u\|_X)$ with $\varphi : \mathbb{R} \to \mathbb{R}_\infty$ proper and even. Then for $v \in X^*$, we have

$$\begin{aligned} F^*(v) &= \sup_{u \in X}\Big\{\langle v, u\rangle - \varphi(\|u\|_X)\Big\} = \sup_{\alpha \geq 0}\left[\sup_{\|u\|_X = \alpha}\Big\{\langle v, u\rangle - \varphi(\alpha)\Big\}\right] \\ &= \sup_{\alpha \geq 0}\Big\{\alpha\, \|v\|_{X^*} - \varphi(\alpha)\Big\} = \sup_{\alpha \in \mathbb{R}}\Big\{\alpha\, \|v\|_{X^*} - \varphi(\alpha)\Big\} = \varphi^*(\|v\|_{X^*}). \end{aligned}$$

   Consider $\varphi(u) = |u|$. Then

$$\varphi^*(y) = \sup_{\alpha \in \mathbb{R}}\Big\{\alpha y - |\alpha|\Big\} = \left\{\begin{array}{ll} 0, & |y| \leq 1 \\ \infty, & \text{else} \end{array}\right\} = \mathbb{I}_{\{|y| \leq 1\}}.$$

   Consider $\varphi(x) = \frac{1}{2}x^2$. Then

$$\varphi^*(y) = \sup_{\alpha \in \mathbb{R}}\Big\{\alpha y - \frac{1}{2}\alpha^2\Big\} = y^2 - \frac{y^2}{2} = \frac{1}{2}y^2.$$

   So in particular, the following correspondences hold for some domain $\Omega \subset \mathbb{R}^n$.

Functional	Fenchel conjugate
$F(u) = \|u\|_{L^1(\Omega)}$	$F^*(v) = \mathbb{I}_{\{\|v\|_{L^\infty(\Omega)} \le 1\}}$
$F(u) = \frac{1}{2}\|u\|^2_{L^2(\Omega)}$	$F^*(v) = \frac{1}{2}\|v\|^2_{L^2(\Omega)}$

2. Indicator function of closed α-balls.
   Let $\alpha > 0$ and $F(u) = \mathbb{I}_{\{\|u\|_X \le \alpha\}}$. Then for $v \in X^*$, it holds

$$\begin{aligned} F^*(v) &= \sup_{u \in X} \left\{ \langle v, u \rangle - \mathbb{I}_{\{\|u\|_X \le \alpha\}} \right\} = \sup_{u \in X,\ \|u\|_X \le \alpha} \langle v, u \rangle \\ &= \alpha \sup_{u \in X,\ \|u\|_X \le 1} \langle v, u \rangle = \alpha \|v\|_{X^*} . \end{aligned}$$

3. Multiplication by positive constants.
   Let $\lambda > 0$ and $\tilde{F} = \lambda F$. Then

$$\begin{aligned} \tilde{F}^*(v) &= \sup_{u \in X} \left\{ \langle v, u \rangle - \lambda F(u) \right\} = \sup_{u \in X} \left\{ \lambda \left\langle \lambda^{-1} v, u \right\rangle - \lambda F(u) \right\} \\ &= \lambda \left(\sup_{u \in X} \left\{ \left\langle \lambda^{-1} v, u \right\rangle - F(u) \right\} \right) = \lambda F^*(\lambda^{-1} v). \end{aligned}$$

4. Translation.
   Let $\tilde{u} \in X$, $\tilde{v} \in X^*$ and $\tilde{F}(u) = F(u + \tilde{u}) + \langle \tilde{v}, u \rangle$. The Fenchel conjugate can be computed by

$$\begin{aligned} \tilde{F}^*(v) &= \sup_{u \in X} \left\{ \langle v, u \rangle - F(u + \tilde{u}) - \langle \tilde{v}, u \rangle \right\} = \sup_{u \in X} \left\{ \langle v - \tilde{v}, u - \tilde{u} \rangle - F(u) \right\} \\ &= \sup_{u \in X} \left\{ \langle v - \tilde{v}, u \rangle - F(u) - \langle v - \tilde{v}, \tilde{u} \rangle \right\} = F^*(v - \tilde{v}) - \langle v - \tilde{v}, \tilde{u} \rangle . \end{aligned}$$

5. Composition with an invertible linear operator $K \in \mathcal{L}(X, Y)$.
   Let Y be a real Banach space, $v \in Y^*$ and $\tilde{F}(u) = F(Ku)$. Then

$$\begin{aligned} \tilde{F}^*(v) &= \sup_{u \in X} \left\{ \langle u, v \rangle - F(Ku) \right\} = \sup_{u \in X} \left\{ \left\langle K^{-1} K u, v \right\rangle - F(Ku) \right\} \\ &= \sup_{w = Ku} \left\{ \left\langle w, \left(K^{-1}\right)^* v \right\rangle - F(w) \right\} = \sup_{w \in Y} \left\{ \left\langle w, \left(K^{-1}\right)^* v \right\rangle - F(w) \right\} \\ &= F^* \left(\left(K^{-1}\right)^* v \right) . \end{aligned}$$

Definition 2.14: Biconjugate

The **biconjugate** $F^{**} : X \to \mathbb{R}_\infty$ of a functional F is defined by

$$F^{**}(u) = \sup_{v \in X^*} \left\{ \langle v, u \rangle - F^*(v) \right\} = \sup_{v \in X^*} \left[\langle v, u \rangle - \left(\sup_{w \in X} \left\{ \langle v, w \rangle - F(w) \right\} \right) \right] .$$

For reflexive spaces X we have $F^{**} = (F^*)^*$. The biconjugate coincides with the functional F, if and only if F is convex and lower semicontinuous. A proof can be found, e.g., in [44, Theorem 5.1].

We now introduce the concept of Fenchel-Rockafellar duality for a specific class of problems. In particular, we now consider functionals which are defined as a sum of a (in a certain sense that will be explained shortly) simple functional and one functional that comprises a linear operator. We aim at exploiting the specific structure of the problem in order to simplify its solution. In the following, we consider the **(primal) problem**

$$\inf_{u\in X}\big\{f_1(u)+f_2(Cu)\big\}, \tag{2.1}$$

for $f_1 : X \to \mathbb{R}_\infty$, $f_2 : Y \to \mathbb{R}_\infty$, both proper, convex and lower semicontinuous and $C \in \mathcal{L}(X,Y)$. Replacing then f_2 by f_2^{**} yields

$$\inf_{u\in X}\sup_{y^*\in Y^*}\big\{f_1(u)+\langle y^*, Cu\rangle - f_2^*(y^*)\big\}. \tag{2.2}$$

This formulation is called the **saddlepoint problem**. Under mild conditions the infimum and supremum can be swapped. Exemplary conditions can be found in [54, Chapter VI]. This yields

$$\begin{aligned}
&\inf_{u\in X}\sup_{y^*\in Y^*}\big\{f_1(u)+\langle y^*, Cu\rangle - f_2^*(y^*)\big\}\\
&=\sup_{y^*\in Y^*}\Big[-\sup_{u\in X}\big\{-f_1(u)+\langle -C^*y^*, u\rangle - f_2^*(y^*)\big\}\Big]\\
&=\sup_{y^*\in Y^*}\big\{-f_1^*(-C^*y^*)-f_2^*(y^*)\big\},
\end{aligned} \tag{2.3}$$

where we used the definition of f_1^* and $\inf F = -\sup(-F)$. The maximization problem in (2.3) is called the **dual problem**. One specific sufficient condition for this exchangeability is given in the following theorem.

Theorem 2.8: Fenchel-Rockafellar duality

Let $f_1 : X \to \mathbb{R}_\infty$ and $f_2 : Y \to \mathbb{R}_\infty$ be proper, convex and lower semicontinuous and $C \in \mathcal{L}(X,Y)$. Suppose the minimization problem (2.1) has a solution u^* and it exists u^0 such that $f_1(u^0) < \infty$, $f_2(Cu^0) < \infty$ and f_2 is continuous at Cu^0. It then holds

$$\inf_{u\in X}\Big\{f_1(u)+f_2(Cu)\Big\} = \sup_{w\in Y^*}\Big\{-f_1^*(-C^*w)-f_2^*(w)\Big\}.$$

In particular, a maximizer $w^* \in Y^*$ exists.

For a proof, see for example [30, Theorem 6.68].

If u^* is a solution of the primal problem and y^* is a solution of the dual problem, then

(u^*, y^*) is a saddle point of the Lagrangian L, i.e.,

$$L(u^*, y) \leq L(u^*, y^*) \leq L(u, y^*)$$

for all $u \in X$ and $y \in Y^*$ with

$$L(u, y) := f_1(u) + \langle y, Cu \rangle - f_2^*(y).$$

Conversely, every saddle point (u^*, y^*) of L defines per definition one solution u^* of the primal and one solution y^* of the dual problem. From the saddle point formulation (2.2) we can derive a popular algorithm, the so-called Primal-Dual Hybrid Gradient method (PDHG), proposed by Pock, Cremers, Bischof and Chambolle in 2009 as a method for minimizing a convex relaxation of the Mumford-Shah functional [115] and by Esser et. al in 2010 [57]. By Fermats theorem (Theorem 2.7), we know that a saddle point (u^*, y^*) fulfills

$$\begin{aligned} 0 &\in \partial(f_1)(u^*) + C^* y^*, \\ 0 &\in \partial(f_2^*)(y^*) - Cu^*. \end{aligned}$$

Simple rearrangements, multiplying by $\sigma > 0$ and adding the identity on each side leads to

$$u^* - \sigma C^* y^* \in (\mathrm{id} + \sigma \partial f_1)(u^*), \tag{2.4}$$

$$y^* + \sigma C \ u^* \in (\mathrm{id} + \sigma \partial f_2^*)(y^*) \tag{2.5}$$

The mapping $(\mathrm{id} + \sigma \partial f)^{-1}$ is called the resolvent or proximal mapping of f. For f proper, convex and lower semicontinuous, the mapping is single-valued and maps to the solution of a minimization problem as stated in Definition 2.15. A proof of the coincidence can be found for example in [30, Lemma 6.134].

Definition 2.15: Proximal mapping

For a proper, convex and lower semicontinuous functional $F : X \to \mathbb{R}_\infty$, the **proximal mapping** $\mathrm{prox}_{\tau F} : X \to X$ is defined by

$$\mathrm{prox}_{\tau F}(u) = \arg\min_{v \in X} \left\{ \frac{1}{2\tau} \|v - u\|_X^2 + F(v) \right\}.$$

A functional F is called **prox-tractable** if the associated proximal mapping is easy to compute and has a closed form.

The proximal mapping is well-defined as the functional minimized is proper, strictly convex and lower semicontinuous and thus a minimizer exists and is unique. From Equations (2.4) and (2.5), we can thus derive the fix-point scheme

$$\begin{aligned} u^{k+1} &= \mathrm{prox}_{\sigma f_1}\left(u^k - \sigma C^* y^k\right), \\ y^{k+1} &= \mathrm{prox}_{\sigma f_2^*}\left(y^k + \sigma C u^{k+1}\right), \end{aligned}$$

which constitutes a proximal descent in the variable u and a proximal ascent in the variable y. It is not clear that such iterations converge, but by introducing an additional over-relaxation step, we can obtain a convergent algorithm. The algorithm (given in Algorithm 1) is generally known as PDHG and was introduced by [57, 115].

Algorithm 1 Primal-dual hybrid gradient method (PDHG)

1: Input: initial pair of primal and dual value u^0, y^0, step size parameters σ, τ, operators C, C^*
2: **for** k=0,1,2,... **do**
3: $\quad u^{k+1} = \operatorname{prox}_{\sigma f_1}\left(u^k - \sigma C^* y^k\right)$
4: $\quad y^{k+1} = \operatorname{prox}_{\tau f_2^*}\left(y^k + \tau C\left(2u^{k+1} - u^k\right)\right)$
5: **end for**

This algorithm is popular in imaging applications as many functionals used in the field are prox-tractable and thus lead to computationally easy schemes for the PDHG algorithm. Some functionals that appear frequently and also in later chapters of this work are listed and their proximal operators are computed in the following Example 2.4, associated calculus rules are stated in Lemma 2.6 (from [30, Lemma 6.136]). Exemplary functions and corresponding proximal operators are depicted in Figure 2.2 to illustrate the concept.

Lemma 2.6: Calculus for resolvents

Let F_1 and F_2 be proper, convex and lower semicontinuous functionals on Hilbert spaces mapping to $\mathbb{R}_\infty$ and let $\sigma > 0$.

1. For α, $\beta > 0$, $\gamma \in \mathbb{R}$ and $F_2(u) = \alpha F_1(\beta u) + \gamma$, it follows that
$$\operatorname{prox}_{\tau F_2}(u) = \beta^{-1}\operatorname{prox}_{\tau\alpha\beta^2 F_1}(\beta u).$$
2. For $\bar{u}$ and w with $F_2(u) = F_1(u + \bar{u}) + \langle u, w\rangle$, it follows that
$$\operatorname{prox}_{\tau F_2}(u) = \operatorname{prox}_{\tau F_1}(u + \bar{u} - \tau w) - \bar{u}.$$
3. For $F_3(u, v) = F_1(u) + F_2(v)$, it holds that
$$\operatorname{prox}_{\tau F_3}(u, v) = \begin{pmatrix} \operatorname{prox}_{\tau F_1}(u) \\ \operatorname{prox}_{\tau F_2}(v) \end{pmatrix}.$$

A proof can be found in [30, Lemma 6.136].
Note that the third rule (a similar statement holds for integrands instead of summands, see the following example, statement 1) can be used to simplify computations for L^p norms. It then suffices to consider the absolute value function and its exponentials.

Example 2.4: Proximal mappings

1. Convex integrands.
Let $X = L^2(\Omega)$ and $\varphi : \mathbb{R}^N \to \mathbb{R}$ be proper, convex and lower semicontinuous

with $\varphi \geq 0$ and either $\varphi(0) = 0$ if Ω has infinite measure or φ bounded from below if Ω has finite measure. Let $F(u) = \int_\Omega \varphi\left(u(x)\right) \mathrm{d}x$. The minimization problem defining the proximal mapping then reads

$$\min_{v \in L^2(\Omega)} \int_\Omega \frac{1}{2\tau} \left|v(x) - u(x)\right|^2 + \tau\varphi\left(v(x)\right) \mathrm{d}x.$$

The subdifferential of F is given by (see [30, Example 6.50] for a derivation)

$$\partial F(u) = \Big\{ w \in L^2(\Omega) \,|\, w(x) \in \partial\varphi\left(u(x)\right) \text{ for almost all } x \in \Omega \Big\}.$$

We can thus define the proximal mapping pointwise by

$$0 \in v^*\left(x\right) - u(x) + \tau\partial\varphi\left(v^*(x)\right) \text{ for almost all } x \in \Omega$$

such that

$$\operatorname{prox}_{\tau F}(u) = \operatorname{prox}_{\tau\varphi} \circ u.$$

2. Absolute value function.
   Let $F : \mathbb{R} \to \mathbb{R}$, $F(u) = |u|$. Then

$$\operatorname{prox}_{\tau F}(u) = \arg\min_{v \in \mathbb{R}} \frac{1}{2\tau} \left|v - u\right|^2 + \left|v\right|.$$

   With $\partial F(v^*) = \begin{cases} \operatorname{sign}(v^*), & v^* \neq 0 \\ [-1, 1], & v^* = 0 \end{cases}$, the optimality condition for the minimizer v^* of the right side is given by

$$0 = \frac{1}{\tau}\left(v^* - u\right) + \partial F(v^*)$$

   which yields

$$\operatorname{prox}_{\tau F}(u) = \begin{cases} u + \tau, & u < -\tau \\ 0, & u \in [-\tau, \tau] \\ u - \tau, & u > \tau. \end{cases}$$

   This mapping is known as soft-thresholding operator.

3. Absolute value plus non-negativity constraint.
   Let $F : \mathbb{R} \to \mathbb{R}$, $F(u) = |u| + \mathbb{I}_{\{u \geq 0\}}$. Then

$$\operatorname{prox}_{\tau F}(u) = \arg\min_{v \geq 0} \frac{1}{2\tau} \left|v - u\right|^2 + \left|v\right|$$

   yields

$$\operatorname{prox}_{\tau F}(u) = \begin{cases} 0, & u \in (-\infty\, , \tau] \\ u - \tau, & u > \tau. \end{cases}$$

4. Absolute value squared.
   Let $F : \mathbb{R} \to \mathbb{R}$, $F(u) = \frac{1}{2}|u|^2$. Then

$$\operatorname{prox}_{\tau F}(u) = \arg\min_{v \in \mathbb{R}} \frac{1}{2\tau}|v-u|^2 + \frac{1}{2\tau}|v|^2 .$$

   F is differentiable with $\partial F(v^*) = v^*$. The optimality condition for the minimizer v^* of the right side is given by

$$0 = \frac{1}{\tau}(v^* - u) + v^*,$$

   which yields

$$\operatorname{prox}_{\tau F}(u) = \frac{u}{1+\tau}.$$

5. Indicator functional.
   Let us first consider closed intervals $I = [a, b]$.
   Let $F : \mathbb{R} \to \mathbb{R}$, $F(u) = \mathbb{I}_{\{[a,b]\}}(u)$ for $a, b \in \mathbb{R}$. Then

$$\operatorname{prox}_{\tau F}(u) = \arg\min_{v \in \mathbb{R}} \frac{1}{2\tau}|v-u|^2 + \mathbb{I}_{\{[a,b]\}}(v) = \arg\min_{v \in [a,b]} \frac{1}{2\tau}|v-u|^2 .$$

   The minimization problem defines the projection operator onto the interval $[a, b]$ such that

$$\operatorname{prox}_{\tau F}(u) = \max\Big(a, \min(b, u)\Big).$$

   Now consider intervals of type $I = (-\infty, b\,]$, the projection operator is then defined by

$$\operatorname{prox}_{\tau F}(u) = \min(u, b) .$$

   For intervals of type $I = [a,\ \infty)$, the projection operator is then defined by

$$\operatorname{prox}_{\tau F}(u) = \max(a, u) .$$

   Note that computing the proximal mapping of indicator functionals as $\mathbb{I}_{\{\|\cdot\|_\infty \leq 1\}}$ simplifies to pointwise projections in each component.

6. Norm functionals.
   Let $F(u) = \varphi(\|u\|_X)$ for $\varphi : \mathbb{R}_+ \to \mathbb{R}_\infty$ proper, convex, lower semicontinuous and increasing. Then

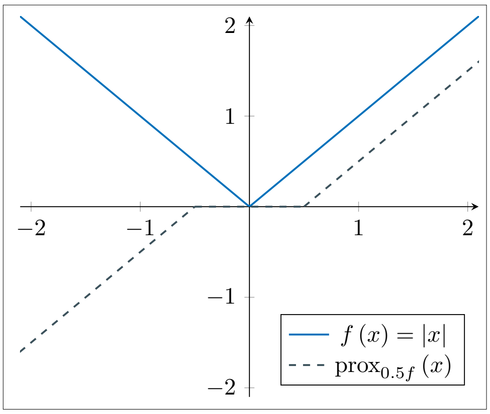
$$\operatorname{prox}_{\tau F}(u) = \arg\min_{v \in X} \frac{1}{2\tau}\|v-u\|_X^2 + \varphi(\|v\|_X) .$$

   For $u = 0$ we see directly $\operatorname{prox}_{\tau F}(u) = 0$. In general, for a minimizer v^* of the right hand side, it holds that

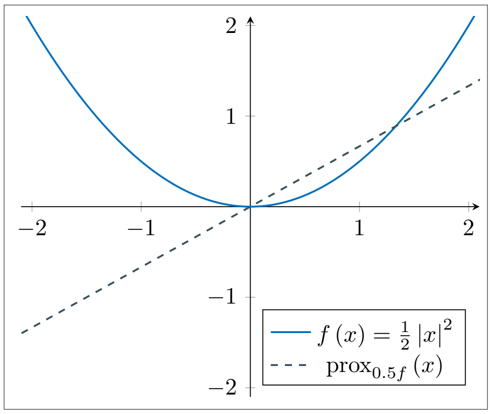
$$0 \in \partial \left(\frac{1}{2\tau} \left\| \cdot - u \right\|_X^2 + \varphi \left(\left\| \cdot \right\|_X \right) \right) (v^*)$$
$$\Leftrightarrow 0 \in \left(\left(\frac{1}{2} \left| \cdot \right|^2 + \tau\varphi(\cdot) \right) \circ \left\| \cdot \right\|_X \right) (v^*).$$

We thus arrive at

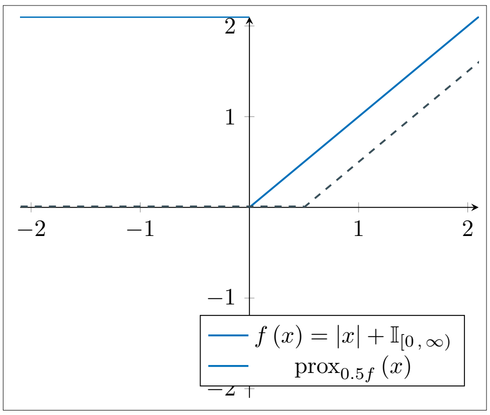
$$\mathrm{prox}_{\tau F}(u) = \mathrm{prox}_{\tau\varphi}(\|u\|_X).$$



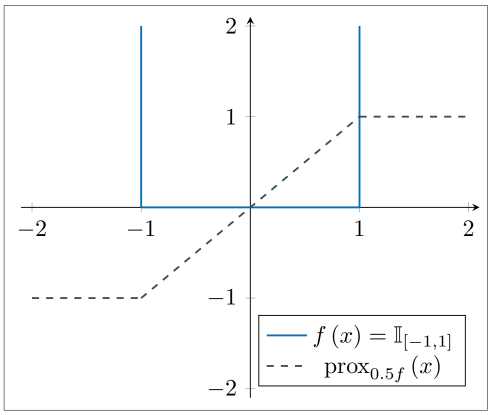
(a) The absolute value function.



(b) The absolute value function squared.



(c) The absolute value plus non-negativity constraint.



(d) An indicator function.

Figure 2.2.: Examples for convex functions and their proximal operators.

2.2.2. Gamma convergence

Later on, we consider a constrained minimization problem that is not solved directly but instead we solve problems incorporating the constraint as a penalty term. We then have to show that the solutions of the unconstrained problems converge to the solution of the constrained problem. In this setting, we need the concept of Γ-convergence as briefly outlined in the following.

Definition 2.16: Equicoercivity

A family $F_n : X \to \mathbb{R}_\infty$ of functionals on a metric space X is called **equicoercive** if and only if it exists a lower semicontinuous and coercive functional $\Psi : X \to \mathbb{R}_\infty$ such that $\Psi \leq F_n$ for all $n \in \mathbb{N}$.

Definition 2.17: Γ-convergence

Let X be a metric space. A functional $F : X \to \mathbb{R}_\infty$ is called the **Γ-limit** of $F_n : X \to \mathbb{R}_\infty$ with respect to the topology of X if

1. For every $u \in X$ and for every sequence $u_n \in X$ with $\lim_{n\to\infty} u_n = u$, it holds that
$$F(u) \leq \liminf_{n\to\infty} F_n(u_n).$$
2. For every $u \in X$ there exists a sequence $u_n \in X$ such that
$$\lim_{n\to\infty} F_n(u_n) = F(u).$$

Remark. There exists a more general definition of Γ-convergence for arbitrary topological spaces that coincides with this definition for this specific case.

Theorem 2.9: Fundamental theorem of Γ-convergence

Let X be a metric space and F_n an equicoercive family of functionals. If $F_n \xrightarrow{\Gamma} F$, i.e., F_n Γ-converges to F in X, then

1. F is coercive.
2. The minimum of F_n converges to the minimum of F, i.e., $\lim_{n\to\infty} \inf_{u\in X} F_n(u) = \min_{u\in X} F(u)$.
3. The minimizers of F_n converge to a minimizer of F, i.e., if $u_n^* \in X$ denotes minimizers of F_n for each n, then a cluster point u^* of u_n^* is a minimizer of F in X.
4. If F has a unique minimum u^*, then u_n^* converges to u^*.

A proof of this theorem can be found in [46, Chapter 7]. More precisely, the first two claims stem from [46, Theorem 7.8.], the third one from [46, Corollary 7.20.] and the fourth one from [46, Corollary 7.24.]. For more details about the concept of Γ-convergence we refer to [26, 46].

2.3. (Dynamic) Inverse problems

The term inverse problem refers to a task where one deduces the cause of an observed outcome. It complements a forward problem where the outcome is observed for a given input. Inverse problems arise naturally in many applications. Examples include image deblurring, where one aims at finding the unpolluted image knowing only a blurry version of it, computed tomography, where one observes the intensity loss of X-rays and reconstructs the tissue of the specimen and magnetic particle imaging, where we measure a change of magnetization and image a tracer distribution. Considering the application of magnetic particle imaging in the following chapters, we need some basic understanding and concepts from inverse problems as stated in this section. For a more comprehensive study, we refer to [55, 86] in English and [118] in German.

Mathematically speaking, for an operator A, and a model

$$Ac = u, \tag{2.6}$$

the forward problem consists of computing the result u for given c, whereas the inverse problem aims at finding c for given u.

Hadamard introduced the notion of an ill-posed problem [69].

Definition 2.18: Well-posedness according to Hadamard

Let $A : X \to Y$ be an operator between two Banach spaces X and Y. The problem (A, X, Y) is called **well-posed** if the following conditions are fulfilled:

1. The equation $Ac = u$ has a solution for every $u \in Y$ (existence),
2. This solution is unique (uniqueness),
3. The inverse operator $A^{-1} : Y \to X$ is continuous (stability).

The problem is called **ill-posed** if any of these conditions is violated.

From a mathematical perspective, the existence condition can be enforced by enlarging the solution space. Uniqueness can be ensured by adding enough additional information to the model, until only one solution exists. The stability condition is most often violated in applications, i.e., the solution does not depend continuously on the data. Thus, small deviations of the input data can lead to large differences in the solution. Even if the data are exact, the numerical solution is most likely unstable as any numerical method has internal errors. Therefore, so-called regularization methods are applied to obtain a stable approximation of the solution. These methods do not minimize the norm discrepancy between Ac and u as this leads to highly oscillating solutions but balance the norm discrepancy and the noise level of the data.

Before we turn to regularization methods, we mention an important family of ill-posed operators, namely compact linear operators.

Definition 2.19: Compact operators

Let X and Y be Banach spaces. The linear operator A is called **compact** if every bounded subset $\mathcal{X} \subset X$ has a pre-compact image $A(\mathcal{X}) \subset Y$, i.e., the closure of the image $A(\mathcal{X})$ is compact.

Theorem 2.10: Ill-posedness of compact operators

Let $A : X \to Y$ be a compact linear operator between infinite dimensional Hilbert spaces X and Y, such that the range of A is infinite. Then problem (2.6) is ill-posed.

See [86] for a proof.
For solving an inverse problem with a compact operator, one could consider a generalized inverse, mapping to an approximate solution. A natural idea to ensure uniqueness is to map to a least-squares solution, i.e., to map to $\bar{c} \in X$ such that $\|Ac - u\|$ is minimum for $c = \bar{c}$. Such an operator mapping u to $\bar{c}$ is called the Moore-Penrose inverse and denoted by $A^\dagger$. Unfortunately, for a compact operator A with infinite dimensional range, the Moore-Penrose inverse $A^\dagger$ defines an unbounded linear operator (e.g., [55, Proposition 2.7.]).
The generalized inverse can, however, be used for classification of the inverse problem. By considering the singular-value decomposition of $A^\dagger$, problems can be categorized mildly ill-posed (if singular values decay with polynomial speed) or severely ill-posed (if singular values decay exponentially). The forward operator in magnetic particle imaging poses a severely ill-posed problem, impeding image reconstruction.
The idea of regularization is the following. To solve (2.6), we try to approximate the generalized inverse solution $c^\dagger = A^\dagger u$ from knowledge of a noisy version u^δ of the data u where $\|u^\delta - u\| \leq \delta$, i.e., we have noise level δ. We look for an approximation that depends continuously on the data u^δ to obtain stability.

Definition 2.20: Regularization

Let $A : X \to Y$ be a bounded linear operator between Banach spaces X and Y, $\alpha_0 \in \mathbb{R}^+$. For every $\alpha \in (0, \alpha_0)$, let $R_\alpha : Y \to X$ be a continuous operator with $R_\alpha 0 = 0$. The family $\{R_\alpha\}$ is called a **regularization** for $A^\dagger$ if there exists a parameter choice rule $\alpha = \alpha(\delta, u^\delta)$ for all $u \in Y$ such that

$$\limsup_{\delta \to 0} \left\{ \left\| R_{\alpha(\delta, u^\delta)} u^\delta - A^\dagger u \right\| \, \middle| \, u^\delta \in Y, \left\| u^\delta - u \right\| \leq \delta \right\} = 0.$$

For the parameter choice rule α it holds that

$$\limsup_{\delta \to 0} \left\{ \alpha\left(\delta, u^\delta\right) \, \middle| \, u^\delta \in Y, \left\| u^\delta - u \right\| \leq \delta \right\} = 0.$$

For a specific α, the pair (R_α, α) is called a **regularization method**.

There are many different possibilities how to define regularization methods, e.g., by applying a filter to the singular value decomposition, by projection and by early stopping of iterative methods. We limit ourselves to Tikhonov regularization which, among others, has a variational characterization. The Tikhonov - regularized solution c_α^δ obtained from u^δ and for parameter choice α is given by the unique minimizer of the **Tikhonov functional**

$$c \mapsto \left\| Ac - u^\delta \right\|_Y^2 + \alpha \left\| c \right\|_X^2 . \tag{2.7}$$

A derivation of different characterizations and the proof for Tikhonov regularization being a regularization method can be seen, e.g., in [55]. We will refer to the Tikhonov functional in (2.7) as classical Tikhonov or standard Tikhonov functional later. However, Tikhonov also considered functionals as

$$c \mapsto \left\| Ac - u^{\delta} \right\|_{Y}^{2} + \alpha \left\| Bc \right\|_{Z}^{2},$$

for an operator $B : X \to Z$ with potentially non-trivial null-space, e.g., a differential operator (see [55, Chapter 8] for details on this setting). Moreover, it is possible to alter the norm used in (2.7) to formally include a-priori knowledge about the solution.

In a static setting as above, there exist numerous theoretical results and numerical methods to solve linear inverse problems. However, there are various applications where the inverse problem is not static but dynamic, e.g., computed tomography imaging suffering from respiratory and cardiac motion, positron emission tomography or functional imaging. Further applications and a general survey on dynamic inverse problems in imaging can be found in [74].

In the dynamic setting, the data as well as the searched-for object are time-dependent, even the forward operator can depend on time, leading to

$$A(t)\, c(\cdot, t) = u(\cdot, t) \quad \text{for } t \in [0, T]. \tag{2.8}$$

We now consider $A(t) : X \to Y$, $c(\cdot, t) \in X$ and $u(\cdot, t) \in Y$ for each t. Including the temporal scale simply as another dimension into existing schemes changes the characteristics of the inverse problem, thus time has to be appropriately incorporated. Still, simply including an additional temporal smoothness prior can be a promising way depending on the specific application, see e.g., [126, 127]. Certainly, it makes the approach more generally applicable compared to more specific schemes. Using a variational framework, reconstruction methods can specifically incorporate motion either by a motion model, see e.g., [34, 35, 50, 61] or as a deformable template [42, 99]. Different classes of ill-posedness for dynamic inverse problems with respect to the Lebesgue-Bochner setting were investigated in [37].

Another large class of inverse problems that we briefly mention is defined by nonlinear inverse problems. In this case, the problem is defined by

$$F(x) = u, \quad x \in \mathrm{dom}(F) \subseteq X, u \in Y,$$

where $F : \mathrm{dom}(F) \subseteq X \to Y$ is a nonlinear operator with domain $\mathrm{dom}(F)$. Later on, we consider a nonlinear problem numerically by splitting it into two linear subproblems. However, there also exists theory on theoretical treatment of nonlinear inverse problems. As this is not within the focus of this work, we refer to [55, 75, 112, 124, 130] for a comprehensive study of different aspects of nonlinear inverse problems.

3. Magnetic Particle Imaging

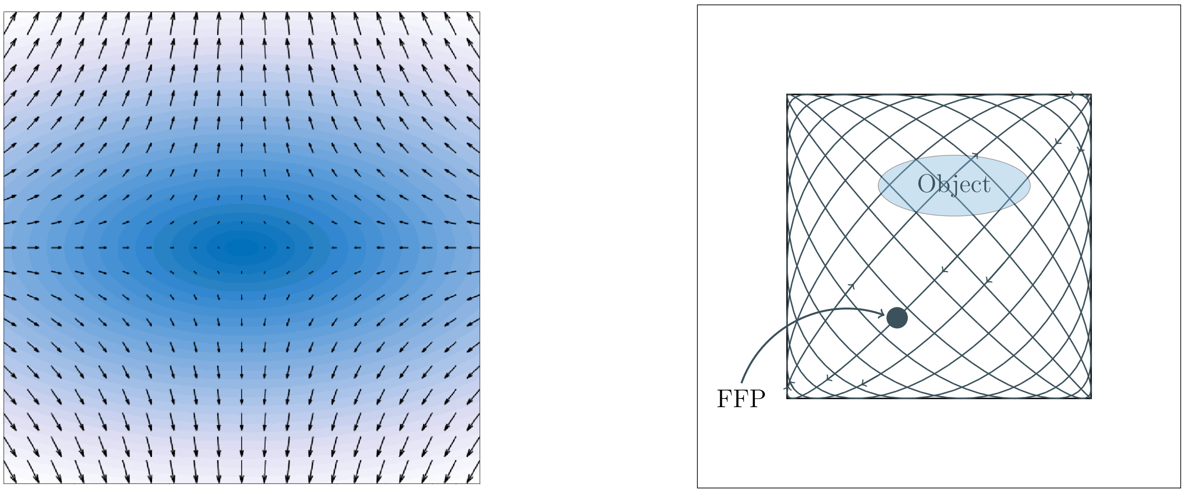
MPI was invented by Bernhard Gleich and Jürgen Weizenecker at Philips Research in Hamburg in the early 2000s. It is thus a relatively new medical imaging method based on the nonlinear magnetization response of magnetic nanoparticles. MPI is a tracer-based method, meaning a contrast agent consisting of Super-Paramagnetic Iron-Oxide Nanoparticles (SPION) is injected into the body prior to MPI measurements and the distribution of the tracer material is reconstructed in the following. There are numerous potential applications like stroke detection [108], cardiovascular imaging [142], instrument tracking during interventions [70] or stem cell monitoring [137, 147]; but to date MPI is still in the preclinical stage.

The main advantages of MPI are its fast 3D acquisition speed and the high spatial resolution. Acquisition times of less than 0.1 s compared to 1 s for CT and 10 s to 30 min for MRI allow for a significantly higher temporal resolution [38]. It has a higher sensitivity in detection of tracers compared to standard methods as MRI or CT [38]. Moreover, no background image of the tissue is obtained such that anatomical background structures do not interfere with the structures of interest. The contrast agent and the magnetic fields used in MPI are not harmful to the human body, whereas other standard methods as CT use ionizing radiation. Another advantage is the proportionality of the generated signal to the amount of SPIONs within the Field-Of-View (FOV).

This chapter is organized as follows. First, we describe the working principle of MPI and derive several properties of the static forward operator in Section 3.1. More particularly, we show compactness and positive definiteness and then derive the main result of this chapter. We show that the static forward operator is non-vanishing over a compact time interval that includes a nonstationary point of the field-free point trajectory. We need this regularity later to apply the proposed joint approach for image reconstruction and simultaneous motion estimation. In Section 3.2, we derive the dynamic MPI forward model and consider the different time scales occurring in the problem. These first two sections are based on [27] and results have first been published therein. We conclude this chapter by Section 3.3, where the image reconstruction process for static MPI is described. We start by deriving the discretization of the continuous problem and then introduce two different reconstruction schemes, one based on the Kaczmarz method and one based on primal-dual splitting. Both algorithmic approaches are applied to synthetic MPI data for a comparison.

3.1. The static forward model

We now briefly introduce MPI and its functionality, for a comprehensive introduction from mathematical as well as engineering perspective, we refer to [91]. An MPI scanner makes use of a time-independent magnetic field (selection field) $H_S : \mathbb{R}^3 \to \mathbb{R}^3$, which has a Field-Free Point (FFP) in the center of the imaging device (see Figure 3.1a). An FFP is the most common geometry, however, a scanner using a field-free line is also considered in the literature [141]. We restrict ourselves to an FFP scanner in this work. The magnetization of magnetic material is saturated by this field everywhere but close to the FFP. This selection field is superimposed by a time-dependent magnetic field (called Drive-Field (DF)) $H_D : [0, T] \to \mathbb{R}^3$, which shifts the FFP in space through the FOV denoted by $\Omega \subset \mathbb{R}^3$. Typically, by the use of sinusoidal excitation functions, the



(a) Selection field.

(b) Lissajous curve.

Figure 3.1.: The selection field H_S has a field-free point (FFP) in the middle (dark blue) and increasing field strength towards the boundaries enabling spatial encoding (a). The FFP is shifted through the region of interest by a time-dependent drive-field. For sinusoidal excitation functions, it follows a Lissajous curve (b).

movement of the FFP follows Lissajous curves (Figure 3.1b) and has highest speed at the center and lowest speed at the edges of the FOV. One cycle of the FFP along that curve is performed within the repetition time T_R. When superimposing the DF, particles in magnetic saturation will not react. Particles in the vicinity of the FFP, however, will experience a strong magnetization and their magnetic moments align with the applied magnetic field. The change of magnetization caused by the FFP movement thus induces a voltage signal by law of induction [91]. This signal is measured by the receive coils of the scanner. As the signal stems only from particles in close vicinity to the FFP, a direct relation between signal and FFP location is established. This spatial encoding can be used for image reconstruction. To mathematically model the signal generation in MPI, we first describe the effective magnetic field $H : \mathbb{R}^3 \times [0, T] \to \mathbb{R}^3$ built by an MPI scanner by the superposition of selection field and drive-field, i.e.,

$$H(x, t) = H_D(t) + H_S(x).$$

We assume that the selection field is a linear gradient field in good approximation. This means there exists a full-rank matrix $G \in \mathbb{R}^{3\times 3}$ such that

$$H_S(x) = Gx.$$

A rank two matrix would produce a field-free line, a geometry that we omit in this work. Note that, in order to keep the area around the FFP contributing to the signal small, the gradient strength has to be sufficiently high. However, the strength applicable in a human scanner is limited due to physiological constraints and must not exceed about $3\,\mathrm{T}/(\mathrm{m}\mu_0)$, limiting the spatial resolution. We can compute the trajectory $x_s : [0,T] \to \mathbb{R}^3$ of the FFP by

$$\begin{aligned} 0 &= H_D(t) + H_S(x) \\ &= H_D(t) + Gx_s(t) \\ \Leftrightarrow x_s(t) &= -G^{-1} H_D(t) . \end{aligned}$$

The time derivative of the effective magnetic field is given by

$$\dot{H}(x,t) = -G\dot{x}_s(t) ,$$

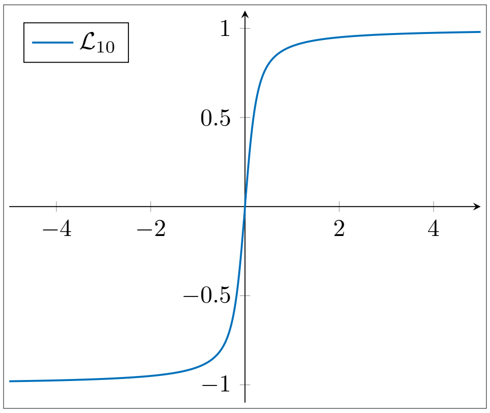
as $H(x,t) = H_D(t) + H_S(x) = -Gx_s(t) + Gx = G(x - x_s(t))$. The static MPI forward operator $A : X \times [0,T] \to Y^L$ for $L \in \mathbb{N}$ receive coil units is defined by

$$\begin{aligned} A_i(c_s,t) = -\int_\Omega c_s(x)\,\mu_0 m_0 R_i^T \Bigg[\left(\frac{\mathcal{L}_\beta'(\|H(x,t)\|_2)}{\|H(x,t)\|_2^2} - \frac{\mathcal{L}_\beta(\|H(x,t)\|_2)}{\|H(x,t)\|_2^3} \right) \\ H(x,t)\,H(x,t)^T + \frac{\mathcal{L}_\beta(\|H(x,t)\|_2)}{\|H(x,t)\|_2} I_3 \Bigg] \dot{H}(x,t)\,\mathrm{d}x, \end{aligned} \tag{3.1}$$

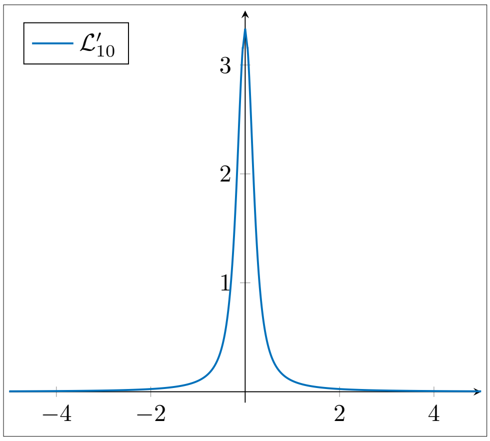
for $i = 1, ..., L$ and $A = (A_i)_i$. Here, $\Omega \subset \mathbb{R}^n$ denotes the FOV, $c_s : \Omega \to \mathbb{R}^+$ denotes the static particle concentration of tracer material, μ_0 describes the permeability constant, m_0 is the absolute of a particle's magnetic moment, $R \in \mathbb{R}^{3\times L}$ denotes the spatially homogeneous receive coil sensitivities (R_i being the i-th column of R), I_3 denotes the identity matrix in $\mathbb{R}^3$ and $\mathcal{L}_\beta$ describes the dilated Langevin function, which is part of the used equilibrium model assumed as the particle magnetization model. Concerning the spaces X and Y, we imagine $X = L^{\hat{p}}\left(0,T; L^l(\Omega)\right)$ for appropriate $\hat{p}$ and $Y = \left(L^2\left(0,T;\bar{Y}\right)\right)^L$ for a reflexive Banach space $\bar{Y}$. This choice will arise naturally to ensure existence of a minimizer in the joint approach for motion estimation and image reconstruction in Chapter 5. Coming back to the equilibrium model, the model assumes a thermal equilibrium of the particles as well as a static magnetic field to be applied and is the most extensively studied magnetization model in MPI. The (dilated) Langevin function $\mathcal{L}_\beta : \mathbb{R} \to \mathbb{R}$ is defined by

$$\mathcal{L}_\beta(z) = \begin{cases} \coth(\beta z) - \frac{1}{\beta z}, & \text{for} \quad z \neq 0 \\ 0, & \text{else,} \end{cases}$$

for a given positive parameter β. The function is illustrated in Figure 3.2a. The magnetization modeled by the Langevin function is saturated for strong applied fields and reacts sensitive to small changes of the applied field within the dynamic region close to zero. The exact value of the dilation parameter models the physics behind the scanner and depends inversely on the Boltzmann constant and the temperature. The derivative



(a) The Langevin function.



(b) The derivative of the Langevin function.

Figure 3.2.: The Langevin function is used in the particle magnetization model. Its nonlinear behavior as well as the saturation property are illustrated by (a). The derivative of the function and its non-negativity is shown in (b).

of the Langevin function, depicted in Figure 3.2b, is given by

$$\mathcal{L}'_\beta(z) = \begin{cases} -\beta \frac{1}{\sinh^2(\beta z)} + \frac{1}{\beta z^2}, & \text{for} \quad z \neq 0 \\ \frac{1}{3}\beta, & \text{else.} \end{cases}$$

Remark. In Figure 3.2, we observe visually continuity of the Langevin function and its derivative and positivity for positive input arguments, in particular for the input arguments in (3.1) for a non-vanishing magnetic field. These properties are needed later and can also be proven analytically. The proof consists of a simple calculation and is left out for brevity.

The forward model for static MPI in n spatial dimensions in the time domain can be similarly expressed by

$$u(t) = \int_\Omega c_s(x) s(x,t) \mathrm{d}x, \tag{3.2}$$

where $u : [0,T] \to \mathbb{R}^L$ denotes the induced signal.
The system function $s : \Omega \times [0,T] \to \mathbb{R}^L$ is described by

$$s_i(x,t) = -\mu_0 R_i^T \frac{\partial \bar{m}(x,t)}{\partial t}, \tag{3.3}$$

for $i = 1, ..., L$ and $\bar{m} : \Omega \times [0,T] \to \mathbb{R}^3$ denotes the mean magnetic moment of the magnetic nanoparticles, i.e.,

$$\bar{m}(x,t) = m_0 \mathcal{L}_\beta \left(\|H(x,t)\|_2\right) \frac{H(x,t)}{\|H(x,t)\|_2}.$$

This shortened notation is introduced to limit the notational burden in the following.

Remark. The MPI signal is T_R-periodic, enabling averaging over several temporal frames to achieve a signal with higher Signal-to-Noise Ratio (SNR). This is a common procedure in MPI as the measured signal usually has a very low SNR caused by the ill-posedness of the forward operator (cf. Lemma 3.1). Note that measurements from at least one full cycle of length T_R are needed to reconstruct an image of the full FOV. Otherwise, the trajectories of the FFP are not closed and some regions might not be scanned at all, whereas in other regions there might be large gaps in the mesh defined by the trajectory. Therefore, a loss of spatial resolution and severe artifacts might occur.

In the following, we denote by $I_{t_s} \subset [0,T]$ a compact interval with nonzero measure that includes the time point $t_s \in [0,T]$. We will specify the interval more precisely when describing the different time scales in MPI. We now analyze the MPI forward operator, starting with a compactness property.

Lemma 3.1: Compactness

Let $I_{t_s} \subset [0,T]$ be compact and with nonzero measure for arbitrary time point $t_s \in [0,T]$. Further, let $\Omega \subset \mathbb{R}^3$ be simply connected and bounded. Moreover, let $\|R_i\|_2 \neq 0$ for $i = 1, ..., 3$, $x_s \in \mathcal{C}^1(I_{t_s})$, and $\dot{x}_s \in \mathcal{C}_b(I_{t_s})$. Assume that $G \in \mathbb{R}^{3\times 3}$ is regular and $c \in L^l(\Omega)$ for arbitrary $l > 1$.
Then the operators $A_{i,t_s} : L^l(\Omega) \to L^2(I_{t_s})$ for $i = 1, \ldots, 3$ with

$$c \mapsto \int_\Omega c(x) s_i(x,t) \mathrm{d}x,$$

for $t \in I_{t_s}$ are compact.

Proof. Let the operator inside the square brackets in (3.1) be denoted by $F(x,t)$ to reduce the notational complexity. The MPI forward operator for the i-th receive coil unit is defined by a linear Fredholm integral equation of first kind with the kernel $s_i : \Omega \times I_{t_s} \to \mathbb{R}$ given by

$$s_i(x,t) = \mu_0 m_0 R_i^T F(x,t) G \dot{x}_s(t) \quad \text{for } t \in I_{t_s}.$$

For a comprehensive introduction to integral equations, we refer to [97]. We now apply [90, Theorem 4.1.] (see Section B.1), to deduce $s_i \in H^0(I_{t_s}; L^\infty(\Omega))$. The assumptions required for the theorem are fulfilled in our setting, as we shortly outline in the following. By $x_s \in \mathcal{C}^1(I_{t_s})$ and $x_s(t) = -G^{-1} H_D(t)$, we know that the DF $H_D(t) \in \mathcal{C}^1(I_{t_s})$ and therefore $H_D(t) \in H^1(I_{t_s})$. Moreover, the selection field $H_S(x) = Gx$ for a full rank matrix G fulfills $H_S \in L^\infty(\Omega)^n$ and the receive coil sensitivity R also fulfills $R \in L^\infty(\Omega)^n$. Thus [90, Equation (4.3)] yields $s_i \in H^0(I_{t_s}; L^\infty(\Omega)) = L^2(I_{t_s}; L^\infty(\Omega))$, i.e., $s_i \in L^2\left(I_{t_s}; L^{l^*}(\Omega)\right)$ for any $1 \leq l^* \leq \infty$.
We now consider s_i as a Hilbert-Schmidt integral operator (see [3] for details) with

$$\|s_i\| := \left(\int_{I_{t_s}} \left(\int_\Omega |s_i(x,t)|^{l^*} \mathrm{d}x \right)^{\frac{2}{l^*}} \mathrm{d}t \right)^{\frac{1}{2}} = \|s_i\|_{L^2\left(I_{t_s}; L^{l^*}(\Omega)\right)} < \infty,$$

with $\frac{1}{l} + \frac{1}{l^*} = 1$ and $1 < l < \infty$. Then by standard results from functional analysis (cf.

[3, Section 5.12]) it follows that the operator

$$\left(A_{i,t_s}c\right)(t) = \int_\Omega c(x)\, s_i(x,t)\,\mathrm{d}x$$

defines a compact operator from $L^l(\Omega)$ to $L^2(I_{t_s})$. □

Remark. In an experimental setup, sinusoidal trajectories are common. As such trajectories are smooth, the assumptions $x_s \in \mathcal{C}^1(I_{t_s})$ and $\dot{x}_s \in \mathcal{C}_b(I_{t_s})$ are not restrictive.

Remark. By Theorem 2.10 we know that compact operators define ill-posed problems. The degree of ill-posedness of the MPI forward operator in terms of decay of the singular values was analyzed in [90]. In a standard setting using the equilibrium magnetization model, trigonometric FFP trajectories and a linear selection field, the singular values decay exponentially yielding a severely ill-posed problem.

Remark. Lemma 3.1 is formulated for $n = 3$ spatial dimensions reflecting the intrinsic 3D nature of MPI. The number of receive coils $L = 3$ is standard for MPI scanners. The result can be transferred to $n = 1$ and $n = 2$ spatial dimensions as follows. We assume that the field-free point $-G^{-1}H_D(t)$ is contained in Ω for all time points $t \in [0,T]$. It is then feasible to restrict to lower dimensional vectors and matrices. We assume that the concentration is Dirac δ-distributed with respect to the orthogonal complement of the lower dimensional affine subspace of $\mathbb{R}^3$. The lower dimensional domain can then be parametrized allowing for a reformulation of the integral [90].

We now prove the regularity assumption on the forward operator needed for Theorem 5.1, i.e., $A_{t_s}\chi_\Omega \neq 0$ for a compact interval I_{t_s} with nonzero measure and arbitrary inner time point $t_s \in I_{t_s}$. In order to do so, we first analyze the positive definiteness on a shifted domain. For fixed t, we define the diffeomorphism $\Phi_t : \Omega \to \Phi_t(\Omega) \subseteq \mathbb{R}^3$ by

$$\Phi_t(\xi) := G^{-1}\xi + x_s(t).$$

Applying the change of variables formula for $x \mapsto \Phi_t(\xi)$ to the forward operator leads to

$$\begin{aligned}
\left(A_{t_s}\chi_\Omega\right)(t) &= \int_\Omega \mu_0 m_0 R^T F(x,t)\, G\dot{x}_s(t)\,\mathrm{d}x \\
&= \int_{\Phi_t^{-1}(\Omega)} \mu_0 m_0 R^T F\left(\Phi_t(\xi), t\right) \left|\det \nabla \Phi_t(\xi)\right| G\dot{x}_s(t)\,\mathrm{d}\xi \\
&= \int_{\Omega_t} \mu_0 m_0 R^T \tilde{F}(\xi) \left|\det\ G^{-1}\right| G\dot{x}_s(t)\,\mathrm{d}\xi,
\end{aligned}$$

with $\Omega_t := \Phi_t^{-1}(\Omega) = G\left(\Omega - x_s(t)\right)$ and

$$\tilde{F}(\xi) := \left(\frac{\mathcal{L}_\beta'\left(\|\xi\|_2\right)}{\|\xi\|_2^2} - \frac{\mathcal{L}_\beta\left(\|\xi\|_2\right)}{\|\xi\|_2^3}\right)\xi\xi^T + \frac{\mathcal{L}_\beta\left(\|\xi\|_2\right)}{\|\xi\|_2} I_3.$$

For this matrix we can verify the following auxiliary lemma.

Lemma 3.2: Positive definiteness

Let $\xi \in \mathbb{R}^3 \setminus \{0\}$ be arbitrary. Then $\tilde{F}(\xi) \in \mathbb{R}^{3\times 3}$ is positive definite.

Proof. We consider $\tilde{F}(\xi)$. Let $\xi \in \mathbb{R}^3 \setminus \{0\}$ and $x \in \mathbb{R}^3$ arbitrary. Consider first

$$x^T \left(\frac{\mathcal{L}_\beta'(\|\xi\|_2)}{\|\xi\|_2^2} - \frac{\mathcal{L}_\beta(\|\xi\|_2)}{\|\xi\|_2^3} \right) \xi\xi^T x = \left(\mathcal{L}_\beta'(\|\xi\|_2) - \frac{\mathcal{L}_\beta(\|\xi\|_2)}{\|\xi\|_2} \right) \frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2}.$$

Moreover,

$$x^T \frac{\mathcal{L}_\beta(\|\xi\|_2)}{\|\xi\|_2} I_3 x = \frac{\mathcal{L}_\beta(\|\xi\|_2)}{\|\xi\|_2} x^T x,$$

yields

$$\begin{aligned} x^T \tilde{F}(\xi)\, x &= \left(\mathcal{L}_\beta'(\|\xi\|_2) - \frac{\mathcal{L}_\beta(\|\xi\|_2)}{\|\xi\|_2} \right) \frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2} + \frac{\mathcal{L}_\beta(\|\xi\|_2)}{\|\xi\|_2} \|x\|_2^2 \\ &= \underbrace{\mathcal{L}_\beta'(\|\xi\|_2)}_{>0} \underbrace{\frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2}}_{\geq 0} + \underbrace{\frac{\mathcal{L}_\beta(\|\xi\|_2)}{\|\xi\|_2}}_{>0} \underbrace{\left(\|x\|_2^2 - \frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2} \right)}_{\geq 0, \text{ as } (x^T\xi)^2 \leq \|\xi\|_2^2 \|x\|_2^2} \geq 0, \end{aligned}$$

where we used the positivity of the Langevin function for positive input arguments, the positivity of the derivative of the Langevin function and the Cauchy-Schwarz inequality. To show positive definiteness, it remains to show that for $x \neq 0$ at least one of the addends is strictly larger than zero. We observe the following.
If $\frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2} = 0$, it follows that $\left(x^T\xi\right)^2 = 0$ and thus $\left(\|x\|_2^2 - \frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2} \right) = \|x\|_2^2 \neq 0$ for $x \neq 0$.
If $\left(\|x\|_2^2 - \frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2} \right) = 0$, then $\frac{\left(x^T\xi\right)^2}{\|\xi\|_2^2} = \|x\|_2^2 \neq 0$ for $x \neq 0$.
We have thus shown

$$x^T \tilde{F}(\xi)\, x > 0 \qquad \text{for } x \neq 0.$$

□

This allows us now to verify the desired regularity property of the forward operator.

Theorem 3.1: Regularity of the MPI forward operator

Let $I_{t_s} \subset [0, T]$ be compact and with nonzero measure for arbitrary inner time point t_s. Assume that the coil sensitivities $R_i \in \mathbb{R}^3$, $i = 1, ..., 3$, fulfill

$$\operatorname{span}\{R_i,\ i = 1, ..., 3\} = \mathbb{R}^3.$$

Moreover, let $x_s \in \mathcal{C}^1(I_{t_s})$ and assume there exists an inner point $t^* \in I_{t_s}$ such that $\|\dot{x}_s(t^*)\|_2 \neq 0$. Let the operator $A_{t_s} : L^l(\Omega) \to L^2(I_{t_s})^3$ be given by $c \mapsto \left(\int_\Omega c(x) s_i(x,t) \mathrm{d}x\right)_{i=1,...,3}$, $t \in I_{t_s}$, with s_i defined according to (3.3).

Then it holds that
$$\|A_{t_s}\chi_\Omega\|_{L^2(I_{t_s})^3} > 0$$
and thus $A_{t_s}\chi_\Omega \neq 0$.

Proof. In the following, we denote $y_s(t) := Gx_s(t)$. As $\{R_i,\ i = 1, ..., 3\}$ forms a basis of $\mathbb{R}^3$, it holds that

$$\forall t \in I_{t_s}: \quad \exists q_i \in \mathbb{R},\ i = 1, ..., 3: \quad \sum_{i=1}^{3} q_i R_i = \dot{y}_s(t).$$

Therefore, it exists $\tilde{q}_i \in \mathcal{C}(I_{t_s}),\ i = 1, ..., 3: \quad \sum_{i=1}^{3} \tilde{q}_i(t) R_i = \dot{y}_s(t)$. Continuity of those functions $\tilde{q}_i$ follows from

$$\underbrace{[R_1\ \ R_2\ \ R_3]}_{=:\bar{T}_R}\tilde{q} = \dot{y}_s \quad \Rightarrow \tilde{q} = \bar{T}_R^{-1}\dot{y}_s,$$

as $\left\|\bar{T}_R^{-1}\right\| < \infty$ and $\dot{y}_s \in \mathcal{C}(I_{t_s})$. For all t with $\dot{y}_s(t) \neq 0$, it holds that

$$\sum_{i=1}^{3} \int_{\Omega_t} \tilde{q}_i(t) R_i^T \tilde{F}(\xi)\dot{y}_s(t)\,\mathrm{d}\xi = \int_{\Omega_t} \underbrace{\dot{y}_s(t)^T \tilde{F}(\xi)\dot{y}_s(t)}_{>0\quad \forall \xi \neq 0}\,\mathrm{d}\xi > 0.$$

For $t^* \in I_{t_s}$ with $\dot{y}_s(t^*) \neq 0$ there exists, by the continuity of $\dot{x}_s$, $\varepsilon > 0$ such that

$$\begin{aligned}
0 &< \int_{(t^*-\varepsilon, t^*+\varepsilon)} \left(\sum_{i=1}^{3} \tilde{q}_i(t) \underbrace{\int_{\Omega_t} R_i^T \tilde{F}(\xi)\dot{y}_s(t)\,\mathrm{d}\xi}_{=:\tilde{u}_i(t)} \right)^2 \mathrm{d}t \\
&= \left\| \sum_{i=1}^{3} \tilde{q}_i \tilde{u}_i \right\|^2_{L^2(t^*-\varepsilon,\ t^*+\varepsilon)} \\
&\leq 3 \sum_{i=1}^{3} \|\tilde{q}_i \tilde{u}_i\|^2_{L^2(t^*-\varepsilon,\ t^*+\varepsilon)} \\
&\leq 3 \sum_{i=1}^{3} \|\tilde{q}_i \tilde{u}_i\|^2_{L^2(I_{t_s})} \\
&\leq 3C_{\tilde{q}} \sum_{i=1}^{3} \|\tilde{u}_i\|^2_{L^2(I_{t_s})} \\
&= 3C_{\tilde{q}} \|\tilde{u}\|^2_{L^2(I_{t_s})},
\end{aligned}$$

where $C_{\tilde{q}} > 0$ denotes an upper bound on $\|\tilde{q}_i\|_{L^2(I_{t_s})}$ for $i = 1, .., 3$, which exists as $\tilde{q}_i \in \mathcal{C}(I_{t_s})$. Denoting now

$$(A_{t_s}\chi_\Omega)(t) = \int_{\Omega_t} \mu_0 m_0 R^T \tilde{F}(\xi) \left|\det\ G^{-1}\right| G\dot{x}_s(t)\,\mathrm{d}\xi$$

$$= \mu_0 m_0 \left|\det\ G^{-1}\right| \widetilde{u}(t)$$

yields the result

$$\|A_{t_s}\chi_\Omega\|_{L^2(I_{t_s})} > 0$$

and thus

$$A_{t_s}\chi_\Omega \neq 0.$$

□

Remark. The assumption span $\{R_i,\ i = 1, ..., 3\} = \mathbb{R}^3$ is not restrictive. In practice, a standard MPI scanner consists of orthogonal receive coils such that $\{R_i, i = 1, ..., 3\}$ forms even an orthogonal basis of $\mathbb{R}^3$.

Remark. By assuming $\|\dot{x}_s(t)\|_2 \neq 0$ for all $t \in [0, T]$, it can also be shown by the same arguments that for the particular choice of $I_{t_s} = \{t_s\}$ the operator $A_{t_s} : L^l(\Omega) \to \mathbb{R}^3$ with $c \mapsto \left(\int_\Omega c(x)s_i(x, t_s)\mathrm{d}x\right)_{i=1,\dots,3}$ fulfills $(A_{t_s}\chi_\Omega) \neq 0$ for all $t_s \in [0, T]$. For our application, this stronger assumption can be fulfilled by choosing an MPI scanner with sinusoidal excitation frequencies for the DF. This choice results in cosine functions for $\dot{x}_s(t)$, which do not share zeros within the timespan $[0, T]$ if the excitation frequencies are chosen accordingly. However, cosine excitation functions lead to sinusoidal derivatives, which share a zero at $t = 0$ and thus do not fulfill the stronger assumption in general.

For the static MPI forward operator we showed the regularity assumption $(A_{t_s}\chi_\Omega) \neq 0$ as well as boundedness of the operators A_{t_s} for desired t_s, which is necessary to prove existence of a minimizer in a joint motion estimation and image reconstruction problem.

3.2. The dynamic forward model

We now extend the MPI forward model to a dynamic tracer distribution $c : \Omega \times [0, T] \to \mathbb{R}^+$, which yields

$$\begin{aligned} u_i(t) &= -\mu_0 R_i^T \frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega c(x, t)\bar{m}(x, t)\mathrm{d}x \\ &= -\mu_0 R_i^T \int_\Omega c(x, t)\frac{\partial}{\partial t}\bar{m}(x, t) + \bar{m}(x, t)\frac{\partial}{\partial t}c(x, t)\mathrm{d}x, \end{aligned} \tag{3.4}$$

for $i = 1, .., L$ assuming again a homogeneous receive coil sensitivity R. The extended model was first mentioned in [63]. The full model can be used for reconstruction, but existing algorithms based on spline interpolation and reconstructing at the temporal resolution of the Analog-to-Digital Converter (ADC) are computationally too expensive to be implemented in practice [28]. The authors of [63] incorporate restrictions to periodic motion in order to reduce the model to the static one.

Moreover, in case of a fully calibrated system function it is still an unsolved problem how to obtain the full dynamic model. We limit ourselves to dynamic distributions whose temporal change provides negligible contributions to the signal such that

$$u_i(t) \approx -\mu_0 R_i^T \int_\Omega c(x, t)\frac{\partial}{\partial t}\bar{m}(x, t)\mathrm{d}x = \int_\Omega c(x, t)s_i(x, t)\mathrm{d}x. \tag{3.5}$$

Considering dynamics in a discrete setting, the interplay of different temporal scales of the measurement process and the dynamics, i.e., of the motion, is an important aspect to take into account. For our application, the important scales are:

1. The scale defined by the ADC. This is the finest temporal scale and defined by the frequency at which measurements are taken (cf. Figure 3.3, ADC sampling). A standard MPI scanner as the Bruker Biospin used for acquisition of the OpenMPIdata [94] delivers 53856 measurements per frame. It is not possible to reconstruct full volumes from just one measurement on this scale.

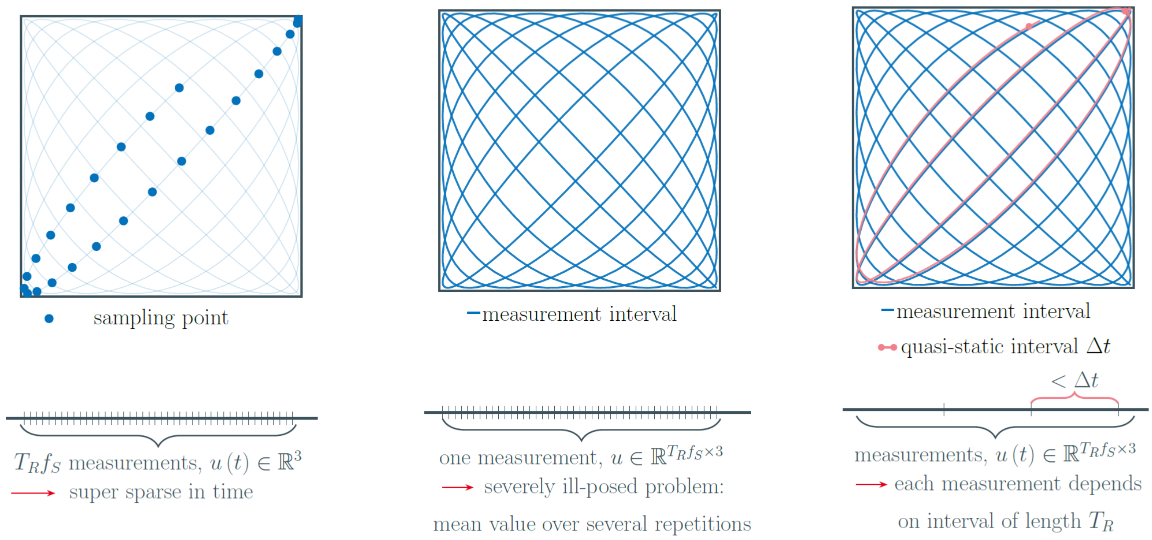
2. The static reconstruction scale defined by the length of one DF cycle (cf. Figure 3.3, Static reconstruction). This time is called repetition time T_R or frame and for a standard 3D Lissajous trajectory it amounts to 21.54 ms, translating to 46 frames per second. Reconstruction of volumes on this scale is possible although averaging over multiple frames is often used in practice.

3. The scale of motion. This scale depends highly on the exact application. Considering cardiovascular imaging, the blood flow velocities differ depending on the type of blood vessels. Maximum velocities inside veins are approximately 10 cm/s, whereas in the aorta it is 12 cm/s and up to 45 cm/s are reached in the arteries [58]. This translates to 2.2 mm/frame, 2.6 mm/frame and 9.7 mm/frame, respectively. To describe the temporal scale corresponding to the motion, we define by Δt the longest duration, for which the tracer distribution remains approximately static, i.e.,

$$c(x,t) \approx c(x,t_s) \qquad \forall t \in [t_s - \Delta t, t_s], \forall x \in \Omega.$$

   Note that if $t_s < \Delta t$, we take advantage of a previous measurement if possible.

MPI is no instantaneous measurement and $c(x, t_s)$ depends on $u(t)$ for $t \in [t_s - T_R, t_s]$ (cf. Figure 3.3, Quasi-static reconstruction).
If $\Delta t \geq T_R$, the motion is called quasi-static and can be handled as static data in good approximation. However, motion artifacts might occur as soon as the motion exceeds one voxel per frame or even earlier, if the spatial grid does not fit the motion very well.
If $\Delta t < T_R$, various problems occur. First, movements during the measurement induce smearing and ghosting artifacts. Second, the FFP trajectory is not closed for a specific state of the tracer such that a loss of spatial resolution is expected. Third, the usual data pre-processing cannot be applied as no frequency selection is possible in order to be able to group and reconstruct data in the time domain. Applying a frequency selection in the Fourier domain would impede application of the Fourier transform to switch to the time domain. This results in a potentially lower SNR. Considering the velocities for cardiovascular imaging mentioned above, this case is of practical relevance as the velocities can easily exceed one voxel per frame.
We restrict ourselves mainly to the assumption $\Delta t = T_R$ in the following, i.e., we assume a quasi-static tracer distribution. This ensures full spatial resolution for each reconstructed image frame but we expect motion artifacts as velocities will exceed one voxel per frame in most settings. The case $\Delta t < T_R$ is considered in Section 6.4, where we use data from subintervals of a frame in order to respect the shorter quasi-static interval.



(a) ADC sampling. **(b)** Static reconstruction. **(c)** Quasi-static reconstruction.

Figure 3.3.: The measurement and data sampling process in MPI. The finest time scale in MPI reconstruction is defined by the sampling rate of the Analog-to-Digital Converter (ADC) sampling measurements at a frequency f_S, see (a). For static reconstruction, we pool all measurements obtained during one drive-field cycle and average over the cycles to gain data with high signal-to-noise ratio (b). For reconstruction of dynamic sequences, we often consider the case of quasi-static reconstruction (c). In this setting, the quasi-static interval Δt (indicated in red) can be smaller than the repetition time T_R (indicated in blue). Still, data from a full drive-field cycle is used to reconstruct the concentration within the field-of-view (cf. (3.6)).

For now, we consider

$$(A_{t_s} c)(t) = \int_\Omega s(x,t)\, c(x,t_s)\, \mathrm{d}x, \quad t \in I_{t_s} := [t_s - T_R, t_s], \tag{3.6}$$

for timepoint $t_s \geq T_R$ to reconstruct $c(x,t_s)$ from u obtained on I_{t_s}. In order to reduce the effects of this approximation in our reconstruction, we jointly reconstruct images and the motion in between time frames such that both tasks will endorse each other.

3.3. Static image reconstruction

For numerical reconstruction of MPI images, we need discretization in space and time. The discretization is derived for the static forward operator. We later on use the same discrete operator as we simplified the dynamic model to the static one. Obviously, we accept an error in the reconstruction by this simplification.
For the time discretization, the finest time scale we can use is defined by the ADC which converts the analog measured signal to a digital one. Remember that reconstructing on this time scale is not possible from one measurement alone, as this would imply reconstructing image information on a 3D volume from a single vector in $\mathbb{R}^3$. However, grouping all information from one DF cycle (see Static reconstruction in Figure 3.3) allows for reconstruction of the concentration within the 3D volume. For static concentrations, re-

construction is usually not performed on this time scale but on data obtained by averaging over several DF cycles.

In MPI, image reconstruction methods can be divided into two groups: model-based approaches and measurement-based approaches. The basic difference is given by the forward operator used for reconstruction. In the model-based approach the particles' physics as well as the measurement process are modeled by physical laws. This aims at being able to obtain a forward operator for different experimental setups rapidly. To achieve this, a magnetization model like the Langevin model presented above is applied to describe the particles' behavior. However, it is very challenging to balance simplicity of the model and its accuracy, and therefore this is an ongoing direction of research. We refer to [87] for an extensive review of existing model-based approaches. Currently, best results are obtained by using measured system matrices, i.e., by using a measurement-based approach. For these approaches, the forward operator is determined prior to the measurements via a time-consuming calibration scan for a given combination of scanner configuration, FFP trajectory and particle type. This is done by moving a delta probe through the FOV and measuring the response signal at each sampling position. The discretization in space is automatically obtained by this calibration.

Typically, we do not consider the forward model in the time domain as in (3.2), but in the Fourier domain, leading to

$$\hat{u}(k) = \int_{\Omega} \hat{s}_k(x) c(x) \mathrm{d}x, \quad k \in \{1, ..., K\},$$

where K denotes the maximum frequency, $\hat{u}(k)$ denotes the Fourier coefficients of the measured voltage u and $\hat{s}_k(x)$ denotes the k-th component of the system function in the Fourier domain. Theoretically, the integration domain is discretized using a regular grid of N points and midpoint rule, yielding the approximate imaging equation

$$\hat{u}(k) = \sum_{n=1}^{N} \hat{s}_{k,n} c_n, \quad k \in \{1, ..., K\},$$

or in matrix-vector form

$$Ac = u, \tag{3.7}$$

where $A \in \mathbb{C}^{K \times N}$ denotes the discrete complex forward operator, $u \in \mathbb{C}^K$ denotes the measured signal and $c \in \mathbb{R}^N$ denotes the discrete concentration. From this approximate equation and the measured signal, we aim at obtaining the spatial distribution of the tracer concentration within the FOV.

The forward operator in MPI is compact and thus ill-posed. Further, the system (3.7) is highly underdetermined such that we need regularization for stable reconstruction.

Numerous different algorithms have been proposed in the context of MPI. Direct solvers as QR decomposition are computationally too expensive [98]. Alternating Direction Method of Multipliers (ADMM) [15, 77, 148], a row-action based method based on sparsity in the wavelet domain [102], a combination of the Kaczmarz method with proximal steps to include a sparsity prior [101], a combination of ADMM and the Kaczmarz method [110] as well as a method based on a deep image prior [52] are, among others, studied

in the literature. Recently, the focus lay on learning-based methods. A physics-driven method based on a deep equilibrium model with learned data consistency was proposed in [68]. In constrast to purely data-driven methods, physical constraints embodied in the system matrix can be obeyed by the learned data consistency. In [7] the authors propose a task agnostic prior trained on an MPI-like data set and integrate it into an ADMM scheme. This prior is combined with the Kaczmarz method in [139], resulting in a significantly shorter run time and high quality results on the considered phantom. Integrating a deep-learning based approach into the post-processing scheme in order to combine several Kaczmarz reconstructed images into one was proposed in [92], obviating the need for manual parameter tuning.
The main challenge is to design a fast and efficient algorithm that can perform online reconstruction but at the same time obtains images with high spatial resolution. The current state-of-the-art reconstruction scheme builds on classical Tikhonov regularization in combination with an L^2-data fidelity term [93, 95, 116, 142] and solves the problem by Kaczmarz iterations. However, an L^1-data fidelity term was shown to have advantages in [89]. A promising technique is the Stochastic Primal-Dual Hybrid Gradient method (SPDHG), the algorithm we use for the image reconstruction subtask later on in this work.
The remainder of the section is organized as follows. We introduce the Kaczmarz algorithm as the most popular reconstruction method used in MPI in Section 3.3.1. Afterwards, we derive the SPDHG algorithm in order to be able to solve a more flexible variational formulation of the problem in Section 3.3.2. Both algorithmic approaches are compared on simulated data in Section 3.3.3.

3.3.1. The Kaczmarz method

The Kaczmarz algorithm solves an optimization problem based on classical Tikhonov regularization in combination with an L^2-data fidelity term [93, 95, 116, 142], i.e., the optimization problem considered is given by

$$\min_c \|Ac - u\|_2^2 + \lambda \|c\|_2^2 , \tag{3.8}$$

for $\lambda > 0$. The iterative algorithm was proposed by Kaczmarz in 1937 [79]. Starting from an initial guess c_0, the problem (3.8) is solved by iterating

$$c_{n+1} = c_n + \lambda \frac{u_i - \langle a_i, c_n \rangle}{\|a_i\|_2^2} \overline{a_i}, \qquad n = 0, 1, ..., \text{ and } i = n \mod K + 1,$$

where a_i denotes the i-th row of the matrix A and K denotes the number of rows of A. The algorithm represents a row-action method, meaning that only one row of the system is used and thus has to be held in memory per iteration. Each iteration performs a projection of the current iterate onto the hyperplane defined by the next row of the equation. The algorithm sweeps through all rows sequentially and then starts over. We consider one full sweep through the rows as one iteration. The number of iterations can be used as an additional regularization parameter. This procedure is called **early stopping**. In case of Lissajous trajectories of the FFP, the rows of the measured MPI system mat-

rix are almost orthogonal leading to fast convergence of the Kaczmarz method [93, 98]. Moreover, the algorithm can easily handle a non-negativity constraint [134]. However, Tikhonov regularization is suboptimal for imaging applications as MPI. In the form used in the MPI literature, the Kaczmarz method does not consider any neighborhood information, i.e., neighbored pixels are assumed to be uncorrelated. Penalizing, e.g., the total variation of the image would more naturally promote image features.
In summary, it can be said that Kaczmarz method is a simple algorithm with fast convergence and low memory consumption, but it solves a problem that is not ideal with respect to the imaging task in MPI reconstruction. More flexibility regarding the regularization term would be beneficial, such that we proposed to use a primal-dual method to solve a fairly general variational formulation in [146].

3.3.2. Stochastic primal-dual hybrid gradient method

This section is based on the just mentioned [146] and summarizes the approach and some main results therein. The paper is joint work with Christina Brandt. We refer to [146] for details. We consider a more general imaging equation, namely

$$\min_c \frac{1}{p} \|Ac - u\|_p^p + \alpha R(Bc) + \beta T(c),$$

where $p \in \{1, 2\}$, $B \in \mathcal{L}(X, Y)$ and R as well as T proper, convex and lower semicontinuous. Moreover, we assume T and the Fenchel conjugate R^* to be prox-tractable. The regularization term T enables integrating prior knowledge of c as for example sparsity or non-negativity but also more complicated prior knowledge that concerns c directly. The term R is mostly used for integrating smoothness assumptions by choosing B as a differential operator. The assumptions on R and T allow for example for L^1- and L^2-norms and also include the Tikhonov regularized formulation in (3.8) with $p = 2$, $\alpha = 0$, $\beta = 2\lambda$ and $T = \|\cdot\|_2$. More particularly, it allows for a TV prior, which is known to accurately estimate discontinuities in images and thus promotes sparsity on the edge set [36, 122, 134]. These characteristics match the features of images in general, and MPI images in particular, very well. For illustration, we consider a specific formulation that was shown to fit MPI images, namely **non-negative fused lasso regularization** in combination with an L^2-data fitting term [134], i.e.,

$$\min_{c \geq 0} \frac{1}{2} \|Ac - u\|_2^2 + \alpha \|\nabla c\|_1 + \beta \|c\|_1 . \tag{3.9}$$

This formulation can be solved by PDHG (see Section 2.2.1 and especially Algorithm 1). In order to do so, the objective function is split into one part containing all functions working directly on c, i.e., $f_1(c) = \mathbb{I}_{\{c \geq 0\}} + \beta \|c\|_1$ and one part containing all functions working on linear operators applied to c, i.e., $f_2(c) = \frac{1}{2} \|Ac - u\|_2^2 + \alpha \|\nabla c\|_1$. The corresponding saddle point problem is derived and solved by a fix point scheme.
In this section, we consider SPDHG, a stochastic version of PDHG, proposed by Chambolle et. al. in 2018 [41]. For this version of the algorithm, only a proper random subset $\mathbb{S}$ of the dual variables is updated in each iteration. Proper in this context means that every dual variable is updated with a positive probability. This approach reduces

the computational costs of evaluating the forward operator and its adjoint per iteration. As a comparable amount of iterations is needed for convergence, this yields faster convergence compared to PDHG. SPDHG has become widely applied in large-scale convex optimization due to its scalability.
When applying SPDHG, the main procedure is the same as for PDHG, but we consider the dual variable split into several variables. For the problem (3.9), the simplest splitting is by defining one dual variable, namely y_1, corresponding to the data term and the system matrix A and one dual variable, namely y_2, corresponding to the total variation prior and the differential operator. The stochastic algorithm for this setting is stated in Algorithm 2.

Algorithm 2 SPDHG for fused lasso regularization and L^2-data fitting

1: Input: $S = [A; \nabla]$, initial values c^0, $y^0 = [y_1^0, y_2^0]^T$, step sizes τ, $\sigma > 0$, regularization parameters α, $\beta > 0$, probability distribution $\tilde{p}$
2: **for** k=0,1,2,... **do**
3: choose data term or TV term update randomly according to probability vector $\tilde{p}$
4: **if** data term update **then**
5: $y_1^{k+1} = \operatorname{prox}_{\sigma\left(\frac{1}{2}\|\cdot\|_2^2\right)}\left(y_1^k + \sigma A c^k\right)$
6: $y_2^{k+1} = y_2^k$
7: **else**
8: $y_2^{k+1} = \operatorname{prox}_{\sigma\left(\mathbb{I}_{\{\|\cdot\|_\infty \le \alpha\}}(\cdot)\right)}\left(y_2^k + \sigma \nabla c^k\right)$
9: $y_1^{k+1} = y_1^k$
10: **end if**
11: $c^{k+1} = \operatorname{prox}_{\tau\left(\mathbb{I}_{\{\cdot \ge 0\}}(\cdot) + \beta\|\cdot\|_1\right)}\left(c^k - \tau p^{-1} S^*\left(2y^{k+1} - y^k\right)\right)$
12: **end for**

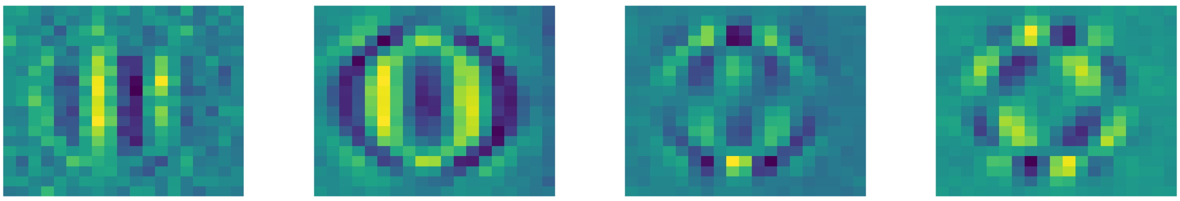
The closed form solutions for the proximal operators can be found in Example 2.4. The extrapolation step in line 11 can be implemented in such a way that only an iterative update is necessary in each iteration, taking into account the dual variable which was updated in that specific iteration [41] and thereby reducing the computational complexity further.
In the following, we give a brief summary of options to adapt SPDHG to MPI reconstruction as outlined in [146].
Using the separability of the MPI forward operator, we can split the dual variable corresponding to the forward operator row-wise into smaller blocks. In this manner, the algorithm can even be transformed into a row-action method. However, it is sufficient to split into data batches small enough to be easily handled in terms of memory limitations. An intuitive idea is to split into three parts corresponding to the three different receive coils of an MPI scanner. This was shown to fasten up convergence of the algorithm. The probability distribution, according to which the updates in the dual step are chosen, can be set freely. We thus choose them data adapted. Therefore, we proposed to link the update probability to the frequency mixing order of the rows. Each row corresponds to a certain frequency component

$$k\Delta f = |m_x f_x + m_y f_y + m_z f_z|\,,$$

where $m_x, m_y, m_z \in \mathbb{Z}$ are called mixing factors and f_x, f_y, f_z are the drive frequencies. The mixing order corresponding to a certain row is then given by

$$m_f = |m_x| + |m_y| + |m_z|\,.$$

For a realistic particle magnetization curve, we can expect a dropping signal with increasing mixing order and thus a low data SNR for high mixing orders. Moreover, the spatial pattern corresponding to certain rows, i.e., the spatial regions which contribute to the signal, can be described by tensor products of Tschebyscheff polynomials. The degrees of the Tschebyscheff polynomials are linked to the mixing orders, a low mixing order is observed in combination with simple spatial patterns. Higher mixing orders are combined with higher polynomial degrees and thus represent rather details than basic information. This relation is illustrated by exemplary spatial patterns corresponding to frequency component k and mixing order m_f in Figure 3.4.



(a) $k = 1633$, $m_f = 5$. **(b)** $k = 2160$, $m_f = 6$. **(c)** $k = 3842$, $m_f = 9$. **(d)** $k = 4337$, $m_f = 10$.

Figure 3.4.: The spatial pattern of a certain row k of the MPI system matrix (corresponding to a specific frequency) depends on the mixing order m_f. The spatial patterns resemble tensor products of Tschebycheff polynomials. This figure includes exemplary patterns observed for the 3D system matrix provided by the OpenMPIData initiative [94]. We depict one slice in the x-y-plane per pattern. Lower mixing orders correspond to more basic patterns, whereas higher mixing orders occur associated with more complex patterns.

Coming back to the probability vector for updates in the dual step in our algorithm, we observed that higher update probabilities for lower mixing orders increase the quality of reconstructed images after a smaller number of iterations. Moreover, it is beneficial to stack real and imaginary parts of the same matrix row into different data batches. Considering the linear operator ∇ used for discretization of the total variation, we found that operator splitting is not needed.

Regarding the convergence properties, convergence in Bregman distance is proven for constant step sizes and under no additional conditions on the functionals in [41]. Under strong convexity assumptions and for very specific step size updates even quadratic convergence rates are shown [41]. In general, appropriate step size rules are the bottleneck for fast convergence. This is already known from deterministic PDHG and also holds for SPDHG. We compared different step size rules empirically, including constant step sizes and adaptive ones based on the variation of gradient directions or the current value of the primal residual in [146]. We found that adaptive step sizes in general perform better than constant ones. This again is a fact known already from PDHG. In [40] a broad class of adaptive step size strategies (including primal-dual balancing step sizes) was introduced and almost-sure convergence of the algorithm was proven. However, for the sake of simplicity we stick to constant step sizes in this work.

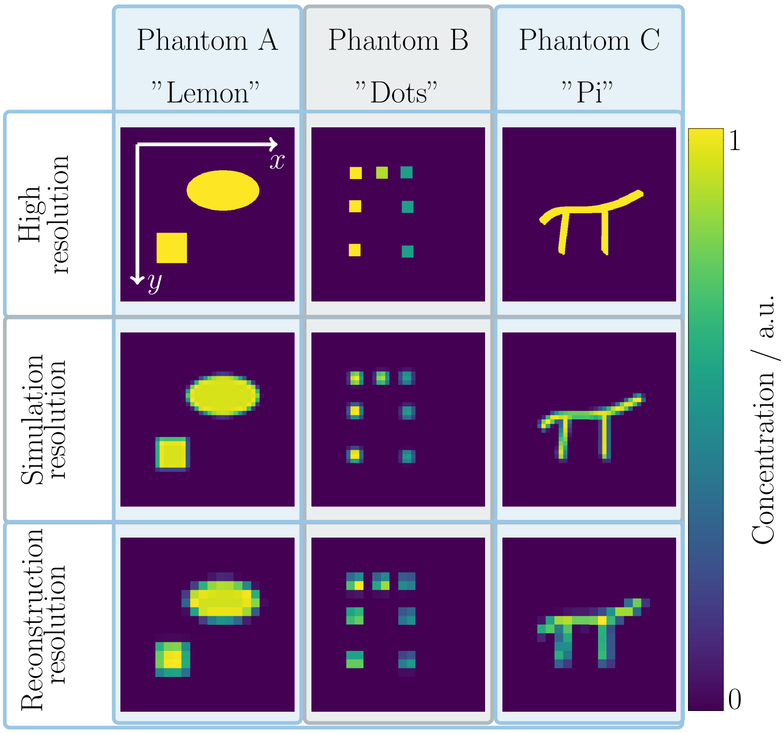


Figure 3.5.: Phantoms for static MPI simulations. The three phantoms are constructed in high resolution and then used for simulation in a slightly lower resolution. Reconstruction is performed on a coarser grid. The "Lemon" phantom features large homogeneous regions, whereas the "Dots" phantom is very sparse and characterized by different concentration levels. The "Pi" phantom is designed in order to resembles a fine vascular structure.

3.3.3. A numerical example on simulated data

In this section, we briefly demonstrate the applicability and strengths and weaknesses of the aforementioned algorithms numerically. In order to be able to compute quality measures of the reconstructed images, we use simulated data for these experiments. The MPI scanner is modeled based on the Bruker Biospin scanner at the University Medical Center Hamburg-Eppendorf. The complete scanner setup can be found in Section A.1.1. We perform simulations under the assumption of static data, i.e., the forward model of MPI is described by (3.1). Three spatial grids are involved to have realistic phantoms and avoid inverse crime. The phantoms are designed on a high resolution grid ($500 \times 500 \times 500$ voxels), simulations are performed on a medium resolution grid ($40 \times 40 \times 40$ voxels) and reconstruction is performed on a low resolution grid ($20 \times 20 \times 20$ voxels). We consider three different phantoms, one consisting of large homogeneous regions (we call this one "Lemon"), one very sparse phantom ("Dots") and one with fine continuous structures resembling fine blood vessels ("Pi"). Each of the phantoms is thus designed having a different potential application of MPI in mind. All three phantoms on the different resolution levels are depicted in Figure 3.5.

Prior to reconstruction, we apply a standard data pre-processing procedure as is also common for measured data. First, we perform a frequency selection. On measured data, one would consider the SNR to select frequencies. For the simulated data, we select in dependence on the maximum mixing order of frequency components and the maximum mixing factor. For this example, we apply a threshold of 10 on the mixing factors and

a threshold of 15 on the mixing orders. Afterwards, we apply row normalization and remove outliers in the data. The removal of outliers can be omitted for algorithms using an L^1-data term as the proximal operator corresponding to the data step then includes shrinkage and thus works as an outlier removal.
In order to find the best reconstruction parameters for each algorithm, we fix applicable constant step sizes and test for each regularization parameter up to differences of order 10, the early stopping parameter for Kaczmarz method is tested on the scale of full iterations. Best parameters are determined based on the Structural Similarity Index Measure (SSIM) of the reconstructed images with respect to the ground truth on the reconstruction scale. An overview on different algorithms, their abbreviations and the used discrepancy and penalty terms is given in Table 3.1. Note that SPDHG can be implemented for L^1- and L^2-data discrepancy terms and we could additionally implement a standard L^2-Tikhonov penalty term, but decided for L^1-, TV- and L^2-norms on the gradient instead.

Table 3.1.: An overview on image reconstruction algorithms as well as discrepancy terms and penalty terms used for each of them. The corresponding abbreviation is of the form Penalty-Discrepancy for all SPDHG schemes and used throughout this chapter for readability. Moreover, we state the number of parameters to tune for each algorithm and list the symbols describing these parameters. A non-negativity term is implemented for each algorithm.

Algorithm	Abbreviation	Discrepancy	Penalty	Number	
Kaczmarz	Kac	L^2	L^2	2	λ, stopping index
SPDHG	FL-L1D	L^1	L^1, TV	2	α, β
SPDHG	FL-L2D	L^2	L^1, TV	2	α, β
SPDHG	L1-L1D	L^1	L^1	1	β
SPDHG	L2Grad-L2D	L^2	$L^2(\nabla)$	1	α

In Figure 3.6 we show reconstructed images obtained by the algorithms stated in Table 3.1 and state the associated SSIM values compared to the ground truth. Each phantom profits from different regularization approaches, which underlines the usefulness of a flexible approach enabling incorporating prior knowledge on the image structure. In a noise-free scenario, all algorithms acquire high-quality results. Only L2Grad-L2D smooths out the concentration too much, we thus consider L^2-regularization on the image gradient as too strong. When reconstructing from noisy data, the images obtained by the Kaczmarz method suffer from lower quality. Moreover, using an L^2-data term in general is problematic. Fused Lasso, i.e., total variation plus L^1-norm (FL) regularization as well as L^1-regularization perform good, but although L1-L1D reaches higher SSIM values, FL-L1D is able to reconstruct homogeneous regions within the image much better.
There is one setting in MPI reconstruction, where using an algorithm with neighborhood respecting regularization is especially useful. As mentioned before, the FOV in MPI is limited due to physiological constraints. More precisely the DF amplitude is limited due to power loss, tissue heating [24] and peripheral nerve stimulation [123, 125]. Therefore, to image larger volumes, it is necessary to combine multiple measurements of smaller regions [2]. This procedure is called **multi-patch measurement**. Applying SPDHG to reconstruct multi-patch MPI data was investigated in [145], which is joint work with M. Boberg and C. Brandt. We showed that the quality of the reconstructed images in a multi-patch scenario is significantly improved by using a regularization that takes into

account neighborhood structures within the images and across the patches' boundaries. In particular, using a TV prior outperforms Kaczmarz reconstructed images in terms of SSIM, Peak-Signal-to-Noise-Ratio (PSNR) and visual inspection on simulated as well as measured data.

3.4. Summary and discussion

In this chapter, we introduced MPI. It is a preclinical imaging method with numerous potential applications. We are interested in reconstruction of dynamic image sequences in order to investigate, e.g., cardiovascular flow inside the human body. After deriving the static forward model of MPI, we showed compactness of the static forward operator and thus saw the ill-posedness of the image reconstruction task. With the help of an auxiliary lemma showing positive definiteness of the forward operator on a shifted domain, we proved Theorem 3.1, the main contribution of this chapter. The theorem claims that on a compact interval with nonzero measure, the forward operator has a positive operator norm on the spatial domain Ω and is thus non-vanishing for non-empty Ω. In the next step, we extended the forward model from static to dynamic tracer distributions. However, using the full dynamic model for image reconstruction is a challenging task on which we do not focus in this work. Therefore, we used the full dynamic model only in order to underline the expectable error in the reconstruction when disregarding the additional part arising from the dynamics. After a comparison of the different time scales in MPI, we decided to limit ourselves to quasi-static reconstruction, i.e., the temporal resolution is limited to one DF cycle. Moreover, we use the static forward operator, although the quasi-static time of the tracer agent might be shorter, for most of the following analysis. However, the topic of subframe reconstruction will be considered in Section 6.4. Having fixed the setting, we addressed the image reconstruction task itself. We briefly introduced the state-of-the-art reconstruction method using the Kaczmarz algorithm. Afterwards, we introduced SPDHG, a stochastic primal-dual splitting approach that yields great flexibility in the reconstruction scheme, as it allows for various different data discrepancy and regularization terms. A comprehensive study on the use of the algorithm in the context of MPI image reconstruction was published in [146]. We summarized the main results, fixed our standard regularization term, namely Fused Lasso regularization and discussed the connection between frequency components in the forward operator and mixing orders as well as Tschebyscheff polynomials. This is particularly interesting, as we use mixing orders in our data pre-processing on simulated data in order to replace the standard SNR thresholding. Applicability of the algorithms was shown on synthetic data for exemplary phantoms.

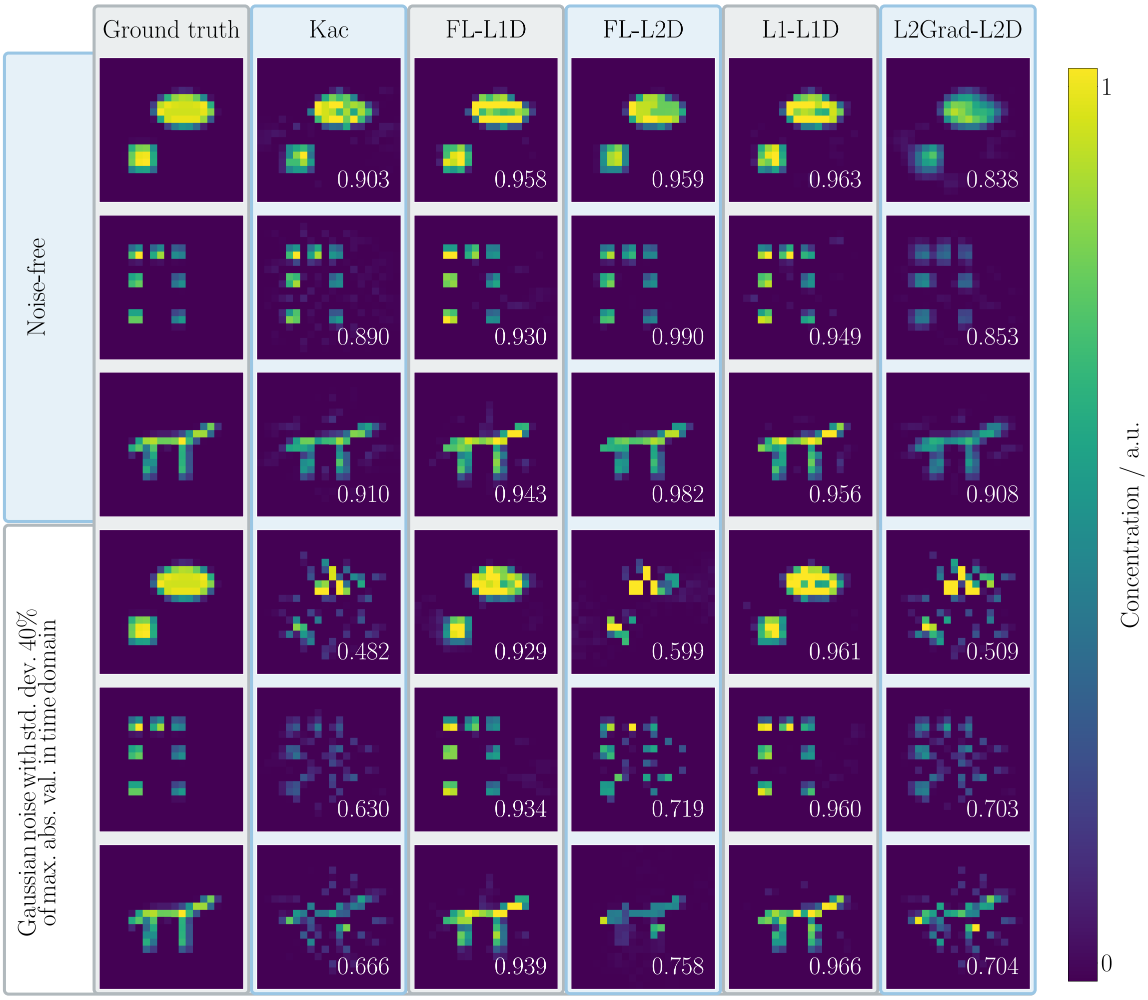


Figure 3.6.: Reconstructed images of simulated static MPI data. The upper three rows depict reconstructions from noise-free data. For the bottom rows gaussian noise with a standard deviation of 40% of the maximum absolute induced voltage was added to the signal in the time domain. This yields noise with a standard deviation of 90% - 130% in the Fourier domain. Each column represents reconstructions by one specific algorithm, the SSIM value is indicated in the bottom right corner for each image. Images reconstructed by Kaczmarz method, the current state-of-the-art method, have high SSIM values in the noise-free setting. However, the phantoms are not recognizable when reconstructed from noisy measurements. In general, the L^2-data discrepancy term struggles differentiating between signal and background noise as we can also observe when considering FL-L2D and L2Grad-L2D algorithms. Reconstructions using an L^1-data fidelity term have significantly higher SSIM values for both Fused-Lasso and L^1-penalty term.

4. Motion Estimation

Motion estimation, i.e., estimating the displacement field between two consecutive images is a key problem in computer vision. There are numerous applications, e.g., studying turbulences in meteorological images, analysis of flows in fluid dynamics and stabilization of videos. In the context of medical imaging, possible applications include correction of, e.g., respiratory motion in lung CT images or cardiac motion in PET images, acquisition of flow patterns in the blood stream or instrument tracking during interventions. In this chapter, our goal is to derive an algorithm to perform motion estimation on MPI image sequences.

Those images are typically very sparse and have an empty background, which poses a very specific task, having other challenges than for example motion estimation on real-world video sequences. Let us consider motion estimation in general.

Two different approaches to motion estimation are distinguished. One can either use the assignment approach, where we search directly for an optimal vector field v such that every voxel x_j at time point $t+1$ can be identified as

$$x_j\,(t+1) = x_i\,(t) + v(x_i, t).$$

This approach conforms to the fact that image sequences are basically image frames sampled along the time axis. It requires identification of feature sets that are to be localized in each frame prior to the assignment task, in order to map the voxels of each feature. Difficulties are caused, e.g., by occlusion and disocclusion of features. Moreover, coping with ambiguity of locally optimal feature matches poses a problem.

The second approach is defined by differential methods, which are based on the assumption that values of a feature mapping g are conserved during motion. This yields a PDE constraint derived from

$$\frac{\partial}{\partial t} g\,(x, t) = 0.$$

Deviations from that constraint are then penalized in a distance functional that could be, e.g., a norm. Displacement fields are obtained by minimizing energy functionals containing such a distance functional. We consider differential methods in this work and discuss different conservation assumptions in Section 4.1.

In the literature, we can again find two different approaches to motion estimation by energy functional minimization. One possibility is using a global energy functional going back to the seminal work [76] proposing the popular Horn-Schunck method, and deriving dense flow fields. On the downside, one has to accept severe fill-in effects in empty image regions and a high sensitivity to noise [13, 62]. Particularly the noise sensitivity might limit the applicability to medical imaging applications, as those images typically suffer from noise artifacts.

The second idea is based on local energy functionals, which are defined in certain regions that might correspond to important features in the images. A seminal work on this branch is [106], introducing the Lucas-Kanade method. Another popular method is the structure tensor approach of Bigün and Granlund [20, 21]. Those methods are typically more robust to noise but, as for the assignment approach, one has to identify important features in a pre-processing step. More recently, a combined local-global method was proposed by [33], which tries to make the best of both approaches. We apply some of their ideas to our motion estimation algorithm in order to improve the sensitivity to noise.

Generally speaking, identifying important features in MPI images should not pose a problem. The images have large empty background regions. Thus, it poses no challenge to distinguish reconstructed tracer material from that background. However, mapping those features by a local method is expected to be a difficult task. The injected tracer agent in MPI might either spread out over time or congregate at certain positions. Matching feature sets from different images might not be possible. We thus consider global methods hoping that the general direction and magnitude of flow can be recovered. The sparse images in MPI pose a challenge to motion estimation in general, as the edge set on which motion estimation algorithms work is extremely sparse.

Comprehensive surveys on optical flow methods and recent developments can be found in [18, 59]. Main challenges in motion estimation are handling motion discontinuities, large displacements and illumination changes and constraining the computational costs. Handling these issues in a unique method still remains an open problem. For our application it is important to especially have a solid performance when facing large displacements as well as illumination changes at a reasonable cost.

This chapter is organized as follows. We start by deriving the two motion models which are used throughout this work, i.e., Optical Flow (OF) and Mass Conservation (MC), from the corresponding conservation laws in Section 4.1. The models are briefly analyzed and compared. Subsequently, we give a brief introduction to algorithmic approaches to motion estimation in Section 4.2 and consider primal-dual schemes to solve the motion estimation problem. We then address some general algorithmic challenges related to motion estimation, more particularly we address multiscale approaches and image warping. Our main contribution in this chapter is the presentation of a multiscale and warping approach for optical flow constrained motion estimation and a multiscale approach for mass conservation constrained motion estimation. Finally, we present some simple numerical examples serving as a test of functionality for the algorithms and discovering the limits of applicability in Section 4.3. Further, we present a new structural prior based on background images in order to improve the stability under noise. Moreover, we briefly analyze the main cost factors and bottlenecks of the algorithms.

4.1. Motion models

A motion model links the image sequence c to the motion field v by defining a conservation assumption. Classical assumptions are conservation of pixel brightness under motion [76] or conservation of mass under motion [47]. Depending on the considered image sequences, one can also assume, e.g., conservation of the pixel gradient magnitude or discontinuities

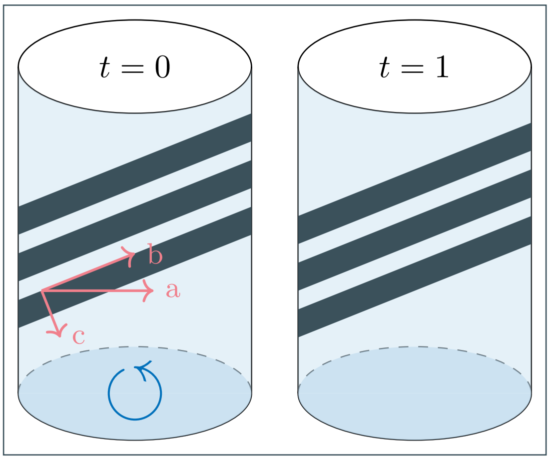


Figure 4.1.: The aperture problem in 2D: only the normal flow, i.e., movement aligned with the image gradient, can be observed but the correct direction cannot be observed (adapted from [10]). The motion component b in direction of the image edges cannot be observed, instead we only observe the component c. Thus, it is not possible to compute the correct value a.

[113]. In this section, we first derive the OF equation from the assumption of constant pixel brightness under motion. Afterwards, we obtain the MC constraint from the assumption of constant mass within the region of interest.

The OF constraint is the most common motion model linking image sequences c over a space time domain $(\Omega \times [0,T]) \subset (\mathbb{R}^3 \times \mathbb{R}^+)$ and motion $v : \Omega \times [0,T] \to \mathbb{R}^3$ in between the image frames. It assumes that voxel intensities stay constant during the motion, i.e.,

$$c\,(x,t) = c\,(x + \delta_t v(x,t), t + \delta_t) \tag{4.1}$$

for δ_t small. By Taylor expansion, one can derive

$$c\,(x + \delta_t v\,(x,t)\,, t + \delta_t) = c\,(x,t) + \delta_t v\,(x,t)\,\nabla c + \delta_t \frac{\partial}{\partial t} c + \mathcal{O}\left(\delta_t^2\right). \tag{4.2}$$

The OF constraint is derived from (4.1) and (4.2) by subtracting $c\,(x,t)$ and dividing by $\delta_t \neq 0$. Moreover, we can neglect $\mathcal{O}\,(\delta_t^2)$ for δ_t small enough, resulting in

$$\frac{\partial}{\partial t} c + v \cdot \nabla c = 0.$$

There are several problems and limitations coming along with the OF constraint. First, the assumption of voxel brightness constancy is violated in many practically relevant applications due to changes in illumination or, in case of MPI images, due to dissemination of tracer agent. As a result, OF cannot be used in case of, e.g., expansion or contraction of an object. Moreover, the model cannot handle occlusion. We also have to assume a high temporal resolution of our image sequence to be able to choose δ_t small enough for the Taylor expansion to hold.

Another main issue is known as the aperture problem in 2D imaging. It means that only movement aligned with the image gradient, i.e., perpendicular to edges, can be observed, c.f. Figure 4.1. Displacement perpendicular to the gradient cannot be observed, as the solution to the OF equation in 2D describes a line and not a single point in space.

In $n = 3$ dimensions, the solution of the OF equation describes not only a line but a plane [14, 132, 133]. The motion estimation problem based on the OF estimation is thus ill-posed in the sense of Hadamard due to the non-uniqueness of the solution and requires regularization for stable computation. While the OF equation is the most common motion model, it has the main drawback that it cannot be used in case of deforming objects. Having in mind our application, this means that nanoparticles are not allowed to gather in one point or scatter over time, but have to stay in the same distribution and are only allowed to move as a whole throughout the measurement.

An alternative motion model overcoming this main drawback is the MC constraint. Instead of assuming constancy of the voxels' brightness, we now assume the mass to remain constant under motion meaning the particles can gather or scatter as long as the total amount of particles stays constant, i.e.,

$$\int_\Omega c\,(x,t)\,\mathrm{d}x = K \quad \forall t \in [0,T]$$

for a constant $K \geq 0$. We furthermore assume that mass cannot be created, annihilated or teleported and thus for a subset $\mathcal{S} \subseteq \Omega$, we assume

$$\frac{d}{dt}\int_\mathcal{S} c\,(x,t)\,\mathrm{d}x = \int_\mathcal{S} \frac{\partial}{\partial t} c\,(x,t)\,\mathrm{d}x,$$

i.e., the increase of mass inside $\mathcal{S}$ must equal the mass which is flowing into the subset through its bounding surface S.

Note that the flow in a point $x \in \Omega$ at time t is given by $c(x,t)v(x,t)$. We can thus determine the flow towards the boundary of $\mathcal{S}$ by

$$\int_{\partial\mathcal{S}} c\,(x,t)\,v\,(x,t)\cdot n\,\mathrm{d}S,$$

where n represents a normal vector to the boundary $\partial\mathcal{S}$ of $\mathcal{S}$. We conclude that

$$\frac{d}{dt}\int_\mathcal{S} c\,(x,t)\,\mathrm{d}x + \int_{\partial\mathcal{S}} c\,(x,t)\,v\,(x,t)\cdot n\,\mathrm{d}S = 0.$$

Applying the divergence theorem yields

$$\int_{\partial\mathcal{S}} c\,(x,t)\,v\,(x,t)\cdot n\,\mathrm{d}S = \int_\mathcal{S} \nabla\cdot\big(c\,(x,t)\,v\,(x,t)\big)\mathrm{d}x,$$

and combining those results leads to

$$\begin{aligned} 0 \quad &= \tfrac{d}{dt}\int_\mathcal{S} c\,(x,t)\,\mathrm{d}x + \int_{\partial\mathcal{S}} c\,(x,t)\,v\,(x,t)\cdot n\,\mathrm{d}S \\ &= \int_\mathcal{S} \frac{\partial}{\partial t} c\,(x,t)\,\mathrm{d}x + \int_\mathcal{S} \nabla\cdot\big(c\,(x,t)\,v\,(x,t)\big)\mathrm{d}x. \end{aligned}$$

Since this holds for arbitrary subsets $\mathcal{S} \subseteq \Omega$, we derive the MC constraint

$$\frac{\partial}{\partial t} c + \nabla\cdot(cv) = 0.$$

The relation between the OF constraint and the MC constraint is shown by

$$\frac{\partial}{\partial t}c + \nabla \cdot (cv) = \frac{\partial}{\partial t}c + c\nabla \cdot (v) + v \cdot \nabla c = 0.$$

For an incompressible flow v, the MC constraint coincides with the OF constraint. Applying motion estimation to dynamic MPI, we cannot assume incompressability as the magnetic nanoparticles are expected to gather or scatter over time. A main drawback of the MC motion model is, again, its disability to cope with occlusion. However, considering MPI images we do not expect occlusion and thus expect the MC motion model to fit the images better than the OF motion model. One important aspect when deciding for a motion model is the complexity of integrating it into numerical schemes. In Section 4.2.3, we derive different schemes for the OF and MC constraint.

4.2. Algorithmic approaches

This section shortly derives classical motion estimation algorithms. We use some of the ideas and concepts to improve the motion estimation algorithm used in succeeding chapters. We first describe the method of Horn and Schunck, then shortly introduce the Lucas-Kanade method and the combined local-global method.

The first method described in the following was proposed by Horn and Schunck in [76] in 1981 for 2D image sequences, but can be easily adapted for 3D image sequences. The method is based on minimizing the global energy functional

$$E_{HS}(c,v) = \int_0^T \int_\Omega \left(\frac{\partial}{\partial t}c + v \cdot \nabla c\right)^2 + \alpha\left(|\nabla v_1|^2 + |\nabla v_2|^2 + |\nabla v_3|^2\right) \mathrm{d}x\mathrm{d}t$$

for an image sequence $c : \Omega \times [0,T] \to \mathbb{R}^+$ and a motion field $v : \Omega \times [0,T] \to \mathbb{R}^3$. The method thus considers the motion estimation problem for the OF constraint embedded into an L^2-norm and with a smooth spatial L^2-regularization on the gradient of the flow field. The corresponding Euler-Lagrange equations are solved by the Jacobi method.

The solution of these equations is unique [128]. Moreover, the solution benefits from the penalty term $\alpha\left(|\nabla v_1|^2 + |\nabla v_2|^2 + |\nabla v_3|^2\right)$ as information is filled in in regions where $|\nabla c| \approx 0$. However, that penalty term reveals also a main drawback of the method: the resulting flow fields are very smooth. The method thus faces problems at discontinuities leading to sub-optimal results in many applications. Therefore, numerous variants of the algorithm, which preserve discontinuities, have been proposed [4, 32, 33]. Brox et. al [32] propose the modified L^1-regularization

$$P\left(|\nabla v|^2\right) = \sqrt{|\nabla v|^2 + \varepsilon}, \tag{4.3}$$

where ε is a parameter for numerical reasons only and set to $\varepsilon = 10^{-3}$. An additional weighting parameter was introduced by Bruhn et. al [33], who use the penalty term

$$P\left(|\nabla v|^2\right) = 2\beta^2\sqrt{1 + |\nabla v|^2/\beta^2}, \tag{4.4}$$

which yields a similar problem for $\beta \approx 1$.

An image-driven regularization approach which suppresses smoothing at or across image boundaries was proposed by Alvarez [4]. As the authors are interested particularly in large displacements, they do not use the linearized OF equation and instead solve the Partial Differential Equation (PDE) system by calculating the asymptotic state via Gauss-Seidel iterations. Although their goal resembles ours when having background information available, their method is computationally too expensive for our application having in mind online motion estimation.

In 1999, the authors of [9] analyzed the L^1-norm for the optical flow constraint and showed advantages compared to a quadratic L^2-norm. An efficient duality-based algorithm for an L^1-norm on the optical flow constraint and additional TV regularization was proposed in [144].

Coming back to the Horn-Schunck algorithm, in summary it can be said that it yields dense (global) flow fields and a smooth flow, which is advantageous in many applications. Drawbacks of the Horn-Schunck algorithm include it being relatively slow and its limitations to small displacements. This is a general issue of algorithms based on the OF equation and will be addressed in Section 4.2.3. As a global method, Horn-Schunck is relatively sensitive to noise.

The Lucas-Kanade method proposed first in [106] assumes spatial constancy of the unknown flow field within some neighborhood. Let $J_\rho(\nabla c) = K_\rho * \left(\nabla c \nabla c^T\right)$, where K_ρ is a Gaussian kernel with standard deviation ρ. Then Lucas-Kanade solves the minimization problem

$$\min_v v^T J_\rho(\nabla c)\, v \tag{4.5}$$

for $v : \Omega \times [0, T] \to \mathbb{R}^3$ constant in a small neighborhood. This local method requires definition of features and neighborhoods beforehand and does not yield dense flow fields.

In 2005, the authors of [33] proposed the Combined Local-Global method (CLG) for OF estimation. The scheme combines the Lucas-Kanade and the Horn-Schunck method to accomplish a method that is both: robust to noise and yielding dense flow fields. It is defined by the minimization problem

$$\min_v \int_\Omega v^T J_\rho(\nabla c)\, v + \alpha \left|\nabla v\right|^2 \mathrm{d}x. \tag{4.6}$$

This problem now takes into account neighborhood information in the data term by smoothing within a small neighborhood and is thus more robust under noise [140]. However, the underlying assumption of locally constant flow fields is not necessarily valid and it certainly does not hold at discontinuities in the flow field.

As we are working on medical images which are typically polluted by noise, we expect the smoothing of the input images to help in obtaining more robust motion estimates. However, local constancy of the flow might or might not hold depending on, e.g., the resolution of the underlying medical images.

The methods presented above solve a PDE system via the corresponding Euler-Lagrange equations. The fixed-point algorithms to solve those equations tend to be computationally expensive. However, they usually achieve high quality motion estimates. In practical applications, we are often more interested in fast methods enabling, at the best of times, online motion estimation.

There are some key features which we expect from our motion estimation algorithm. We aim at a global method yielding dense flow fields, robustness to noise, a low computation time, flexibility regarding the regularization term enabling smooth and discontinuous flow estimation and flexibility regarding the motion model enabling estimation of compressible and incompressible flows. Therefore, a motion estimation algorithm that is expected to meet those criteria is presented in the following. We consider another class of algorithms to solve energy functionals arising from the differential approach to motion estimation. Instead of using fixed-point schemes on the Euler-Lagrange equations, we now derive the saddle point formulation of the minimization problems related to the energy functionals and solve them by a primal-dual algorithm.

4.2.1. PDHG for the optical flow constraint

Consider now the problem

$$\min_v \int_0^T \beta S\left(v(\cdot,t)\right) + \gamma \left\| \frac{\partial}{\partial t} c(\cdot,t) + \nabla c(\cdot,t) \cdot v(\cdot,t) \right\|_{L^1(\Omega)}^1 \mathrm{d}t, \tag{4.7}$$

in order to determine $v : \Omega \times [0,T] \to \mathbb{R}^n$, $v \in L^q(0,T; BV(\Omega)^n)$, $q > 1$ for a regularization term $S : BV(\Omega)^n \to \mathbb{R}$. The input image sequence c is given and fixed, as well as the regularization parameter $\beta > 0$ and the weighting term γ. Note that $\gamma = 1$ is a standard choice when considering the motion estimation problem on its own. However, we need the option to adapt the parameter in the joint approach later on. The L^1-norm in the data discrepancy term in (4.7) can be easily replaced by an L^2-norm. Using the L^2-norm yields an optimization problem that is closer to the one considered by Horn-Schunck and Lucas-Kanade, but aims at satisfying the motion model in an averaged sense. The L^1-norm on the other hand accepts outliers that do not fulfill the motion model as long as it is fulfilled in most points and is thus more robust. The regularization functional S is assumed to be proper, lower semicontinuous and convex as well as prox-tractable. For our application, we usually consider, e.g., TV or L^2-regularization of the gradient, i.e.,

$$\begin{aligned} S_1\big(v(\cdot,t)\big) &= \|\nabla v(\cdot,t)\|_{L^1(\Omega)^n} \qquad \text{or} \\ S_2\big(v(\cdot,t)\big) &= \frac{1}{2}\|\nabla v(\cdot,t)\|_{L^2(\Omega)^n}^2 . \end{aligned} \tag{4.8}$$

However, it is also possible to choose standard Tikhonov regularization on v if the flow field is not assumed to be very smooth, i.e.,

$$S_3\big(v(\cdot,t)\big) = \frac{1}{2}\|v(\cdot,t)\|_{L^2(\Omega)^n}^2 .$$

We restrict ourselves to the exemplary case of an L^1-data discrepancy term and penalty term S_1 in order to derive the update steps for our algorithm. We apply primal-dual splitting with primal functional f_1 and f_2 to be dualized defined by

$$\begin{aligned} f_1(v) &= \int_0^T \gamma \left\| \frac{\partial}{\partial t} c(\cdot,t) + \nabla c(\cdot,t) \cdot v(\cdot,t) \right\|_{L^1(\Omega)}^1 \mathrm{d}t, \\ f_2\big(Cv\big) &= \int_0^T \beta \left\| \nabla v(\cdot,t) \right\|_{L^1(\Omega)^n} \mathrm{d}t. \end{aligned}$$

The linear operator C is defined by

$$C = \begin{pmatrix} \nabla & 0 & 0 \\ 0 & \nabla & 0 \\ 0 & 0 & \nabla \end{pmatrix},$$

leading to the adjoint operator

$$C^* = \begin{pmatrix} -\mathrm{div} & 0 & 0 \\ 0 & -\mathrm{div} & 0 \\ 0 & 0 & -\mathrm{div} \end{pmatrix},$$

such that we can formulate the primal problem as

$$\min_v f_1(v) + f_2(Cv).$$

The dual functional corresponding to f_2 is given by

$$f_2^*(y_1, y_2, y_3) = \int_0^T \beta \left(\mathbb{I}_{\left\{\|y_1\|_\infty \le \beta\right\}} + \mathbb{I}_{\left\{\|y_2\|_\infty \le \beta\right\}} + \mathbb{I}_{\left\{\|y_3\|_\infty \le \beta\right\}} \right) \mathrm{d}t,$$

as derived in Example 2.3. The saddle point formulation (see (2.2)) of the motion estimation problem reads

$$\begin{aligned} \min_v \max_y \int_0^T \gamma \left\| \frac{\partial}{\partial t} c(\cdot,t) + \nabla c(\cdot,t) \cdot v(\cdot,t) \right\|_{L^1(\Omega)}^1 &+ \langle y, Cv(\cdot,t) \rangle \\ - \beta \left(\mathbb{I}_{\left\{\|y_1\|_\infty \le \beta\right\}} + \mathbb{I}_{\left\{\|y_2\|_\infty \le \beta\right\}} + \mathbb{I}_{\left\{\|y_3\|_\infty \le \beta\right\}} \right) \mathrm{d}t. \end{aligned}$$

Remark. Let us briefly consider existence of solutions to the motion estimation problem. By Theorem 2.6, solutions exist for reflexive Banach spaces and a functional that is bounded from below, weakly lower semicontinuous and coercive. The energy functional fulfills these assumptions as it is convex and lower semicontinuous, which is equivalent to it being weakly lower semicontinuous, such that a minimizer exists, e.g., on $L^p(0,T;L^s(\Omega))$ for $1 < p, s < \infty$. A similar result holds for the dual of a separable normed space if we assume a weak-* lower semicontinuous functional. We can consider $L^q(0,T;BV(\Omega))$ for $1 < q < \infty$ as the dual of a separable Banach space. By the additional constraint $\|v(\cdot,t)\|_{L^\infty(\Omega)} \le k_\infty$ for each $t \in [0,T]$ for a constant k_∞ we can enforce boundedness of sublevel sets with respect to the weak-* topology. More particularly, the sublevel sets are closed (w.r.t. the weak-* topology) as the energy functional is continuous and bounded in norm due to the additional constraint. Therefore, the functional is weak-*

lower semicontinuous and minimizers exist. Moreover, the assumptions for Theorem 2.8 are fulfilled such that Fenchel-Rockafellar duality holds and the saddle point problem is equivalent to the primal problem.

4.2.2. PDHG for the mass conservation constraint

Consider now the motion estimation problem given by

$$\min_v \int_0^T \frac{\beta}{2} \|\nabla v(\cdot,t)\|^2_{L^2(\Omega)^n} + \gamma \left\| \frac{\partial}{\partial t} c(\cdot,t) + \nabla \cdot \big(c(\cdot,t)\cdot v(\cdot,t)\big) \right\|_{L^1(\Omega)} \mathrm{d}t. \tag{4.9}$$

The algorithm can again also handle an L^2-norm as data discrepancy term and penalty terms S_1, S_2 and S_3, respectively. We confine ourselves to the specific case in (4.9) in order to show how to derive the corresponding algorithmic updates by this example. Fenchel conjugates and proximal operators needed for different choices are stated in Example 2.3 and Example 2.4. The problem is solved by primal-dual splitting with

$$\begin{aligned} f_1(v) &= 0, \\ f_2(Cv) &= \int_0^T \frac{\beta}{2} \|\nabla v(\cdot,t)\|^2_{L^2(\Omega)^n} + \gamma \left\| \frac{\partial}{\partial t} c(\cdot,t) + \nabla \cdot \big(c(\cdot,t)\cdot v(\cdot,t)\big) \right\|_{L^1(\Omega)} \mathrm{d}t \end{aligned}$$

with the linear operator

$$C = \begin{pmatrix} \nabla & 0 & 0 \\ 0 & \nabla & 0 \\ 0 & 0 & \nabla \\ \partial_x c & \partial_y c & \partial_z c \end{pmatrix}$$

and its corresponding adjoint operator

$$C^* = \begin{pmatrix} -\mathrm{div} & 0 & 0 & -c\partial_x \\ 0 & -\mathrm{div} & 0 & -c\partial_y \\ 0 & 0 & -\mathrm{div} & -c\partial_z \end{pmatrix}.$$

The primal problem

$$\min_v f_1(v) + f_2(Cv)$$

can thus be transferred to the saddle point problem

$$\begin{aligned} \min_v \max_y \int_0^T & \langle y, Cv(\cdot,t)\rangle - \frac{1}{2\beta}\left(\|y_1\|_2^2 + \|y_2\|_2^2 + \|y_3\|_2^2\right) \\ & - \gamma \left(\mathbb{I}_{\{\|y_4\|_\infty \le \gamma\}} - \langle c_t, y_4\rangle\right) \mathrm{d}t, \end{aligned}$$

using the Fenchel conjugate f_2^* of f_2 defined by

$$f_2^*(y) = \int_0^T \frac{1}{2\beta}\left(\|y_1\|_2^2 + \|y_2\|_2^2 + \|y_3\|_2^2\right) + \gamma \left(\mathbb{I}_{\{\|y_4\|_\infty \le \gamma\}} - \langle c_t, y_4\rangle\right) \mathrm{d}t.$$

Remark. Existence of solutions as well as applicability of Fenchel-Rockafellar duality can be seen analogously to the OF case.

Using the above notations, the motion estimation algorithm with extrapolation in the dual variable is given by Algorithm 3. The proximal operators corresponding to the functionals f_1 and f_2^* have been derived in Example 2.4.

Algorithm 3 PDHG algorithm for motion estimation

Input: initial values v^0, y^0, θ, stepsize parameters σ, τ, operators C, C^*
for k=0,1,2,... **do**
    $v^{k+1} = \mathrm{prox}_{\tau f_1}\left(v^k - \tau C^* \tilde{y}^k\right)$
    $y^{k+1} = \mathrm{prox}_{\sigma f_2^*}\left(y^k + \sigma C v^{k+1}\right)$
    $\tilde{y}^{k+1} = y^{k+1} + \theta\left(y^{k+1} - y^k\right)$
end for

4.2.3. Algorithmic details: Operator discretization, multiscale and warping

As proposed by [35], we use forward and backward differences for discretization of the differential operator and its adjoint, respectively, for the spatial regularization terms. Within the motion constraint, we use central differences, which are self-adjoint, for the spatial derivatives and forward differences for the time derivatives to obtain a stable scheme.

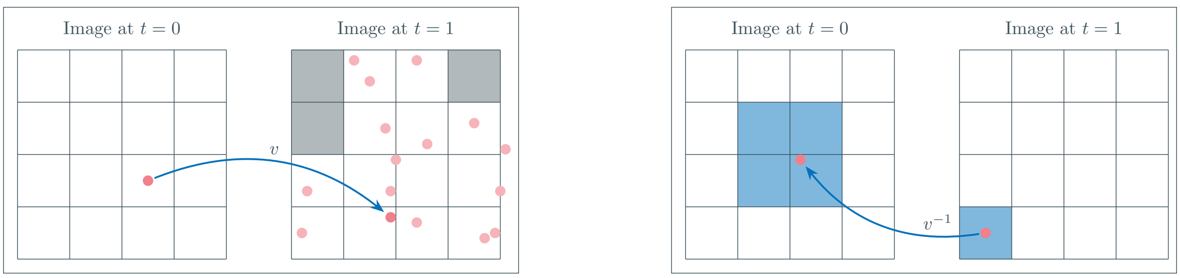
Multiscale plus warping for the optical flow constraint

As the linearized OF constraint is derived by a Taylor approximation, we note that it will hold only for motion of small absolute value. Moreover, the maximum absolute value of motion estimates is limited by the discretization of the differential operators which also limit the detection range in which pixels can be compared. This extends the problem to handle large displacements to the case of the MC constraint. To deal with larger displacements between two consecutive frames, a coarse-to-fine strategy is algorithmically used. There are two different basic approaches to the multiscale strategy. One either computes only motion increments on each scale and warps images in between the different scales [32] or one uses the results from coarser scales for initialization but then updates the whole value in each iteration [111]. The choice also depends on the motion model used as we will see, such that we combine motion increments and warping with the OF motion model and initialization by coarser results with the MC constraint. The author of [51] proposed a multiscale and warping scheme for variational large scale motion estimation based on the OF constraint. They use an upscaled result from coarser scales as initial value on the next finer scale and perform several refinements on each scale applying warping in between.

We now consider the approach used in this thesis. A coarse-to-fine strategy is a multi-resolution approach to the motion estimation. The input image sequence is iteratively down-sampled in the spatial dimensions in order to obtain a coarser version. On the

coarsest scale, the motion considered is potentially small enough to be accurately estimated and moreover, the computational burden on coarser grids is lower. The result from a coarse scale is then prolongated to the next finer scale and either used for warping the images of the next scale or as initial value for the motion estimate on the next scale.
Let $z = 0, ..., Z$, $Z \in \mathbb{N}$ denote the different scales in the coarse-to-fine scheme. By c_0 we denote the image on the coarsest scale and $c_Z \equiv c$ denotes the image on the finest scale, i.e., on the original grid. The coarser images are obtained by applying a down-sampling factor η^{Z-z} to the original image. If $c \equiv c_2$ is for example of size 2^m in each spatial dimension, then for a down-sampling factor of $\eta = 2$, the image c_1 on scale 1 is of size $2^m/2 = 2^{m-1}$ in each direction and c_0 on scale 0 is of size $2^m/2^2 = 2^{m-2}$.
In case of the OF motion model, the coarse-to-fine strategy can be combined with warping.

Remark. Note that image warping can be implemented by a forward mapping or a reverse mapping. When using a forward mapping, we iterate over the pixels of the source image and compute the new position of each pixel in the destination image. The result is then a floating point and we use either nearest neighbor or linear interpolation (or any other) to re-grid the image. However, many source pixels can map to the same destination and some pixels in the destination image might not be covered at all. For this reason, warping is usually defined as a reverse mapping. We then iterate over the pixels of the destination image and find the origin in the source image. The gray value to be transported to the destination image is then again determined by bilinear interpolation (or any other method). The comparison of warping methods is illustrated in Figure 4.2. Reverse warping thus in a sense fulfills the gray value constancy assumption and can be combined with the OF constraint, as that constraint will lead to invertible flow fields. However, in case of MC as motion model, there might exist sources and sinks and the flow field is then not invertible. We thus cannot perform reverse warping in general.



(a) Forward warping. **(b)** Reverse warping.

Figure 4.2.: Forward warping versus reverse warping. Forward warping iterates over the input image at $t = 0$ and computes the pixels position in the next image frame at $t = 1$. Pixels can be uncovered (e.g., the gray pixels) or hit by more than one source pixel. Reverse warping iterates over the pixels of the destination image at $t = 1$ and finds the corresponding position in the source image at $t = 0$. The associated gray value is determined by means of bilinear interpolation.

It was shown that the warping technique implements the nonlinearized OF equation [32]. Combining multiscale and warping means that a motion estimation algorithm is applied on each scale, beginning from the coarsest one. The estimate from the previous scale is prolongated to the next finer one and then used to warp the images accordingly. Afterwards, a motion estimation algorithm is applied on the next scale. The process is illustrated for the case of an input image sequence consisting of two time steps in Figure 4.3. Algorithmically, one has to consider that on each scale only optimal increments are searched-for, whereas an initial value is already given. Moreover, we have to consider the additional warping operator. The multiscale and warping algorithm is presented in Algorithm 4.

Algorithm 4 Motion estimation with multiscale and warping

```
Input: image c = c_Z, initial motion field guess v = v_Z, number of scales Z, down-
  sampling factor η
for z=Z-1,...,0 do
    c_z = imresize (η, c_{z+1})   % build image pyramid
end for
for z=0,..,Z do
    v_z = imresize (η^{Z-z}, v_Z) − Σ_{k<z} imresize (η^{k-z}, v_k)   % initial increments
end for
for z=0,...,Z-1 do
    v_z = detect motion(c_z, initial value v_z, priors v_k for k < z)
    v_z = imresize (η^{-1}, v_z)
    c_{z+1} = warp images(c_{z+1}, v_z)
end for
v_Z = detect motion(c_Z, initial value v_Z, priors v_k for k < Z)
v = v_Z + Σ_{z=0}^{Z-1} imresize (η^{Z-z-1}, v_z)
```

The PDHG algorithm (for OF motion estimation) now applied on each scale is only slightly affected by the transition to a multiscale plus warping scheme. In this case, the formulation to be solved on scale z is given by

$$\min_{v_z} \int_0^T \beta S\left(v_z(\cdot,t) + \sum_{k<z} v_k(\cdot,t)\right) + \gamma \left\| \frac{\partial}{\partial t}\Big(\mathcal{W}_z c_z(\cdot,t), c_z(\cdot,t)\Big) + \nabla\Big(\mathcal{W}_z c_z(\cdot,t)\Big)\cdot v_z(\cdot,t) \right\|_{L^1(\Omega)}^1 \mathrm{d}t,$$

where the operator $\mathcal{W}_z$ denotes a warping operator with respect to the previous (prolongated) motion estimate v_{z-1} and $\frac{\partial}{\partial t}\big(\mathcal{W}_z c_z(\cdot,t), c_z(\cdot,t)\big)$ describes the difference between the warped images at the second time step and the original images at the first time step. However, we do not consider $\mathcal{W}_z$ as part of the linear operator in the Fenchel-Rockafellar setting, but consider $\mathcal{W}_z c_z$ as input argument to our algorithm on each scale. Thus, we have a change only in f_2, in particular the previously applied motion has to be taken into account when applying regularization such that

$$\tilde{f}_2(Cv_z) = \int_0^T \frac{\beta}{2} \left\|\nabla\big(v_z(\cdot,t) + v_{\text{prev}}\big)\right\|_{L^2(\Omega)^n}^2 \mathrm{d}t = f_2(Cv_z + Cv_{\text{prev}}),$$

where v_{prev} denotes all previously applied motion estimates prolongated to scale z. This leads to

$$\tilde{f}_2^*(y) = f_2^*(y) - \langle y, Cv_{\text{prev}}\rangle\,,$$

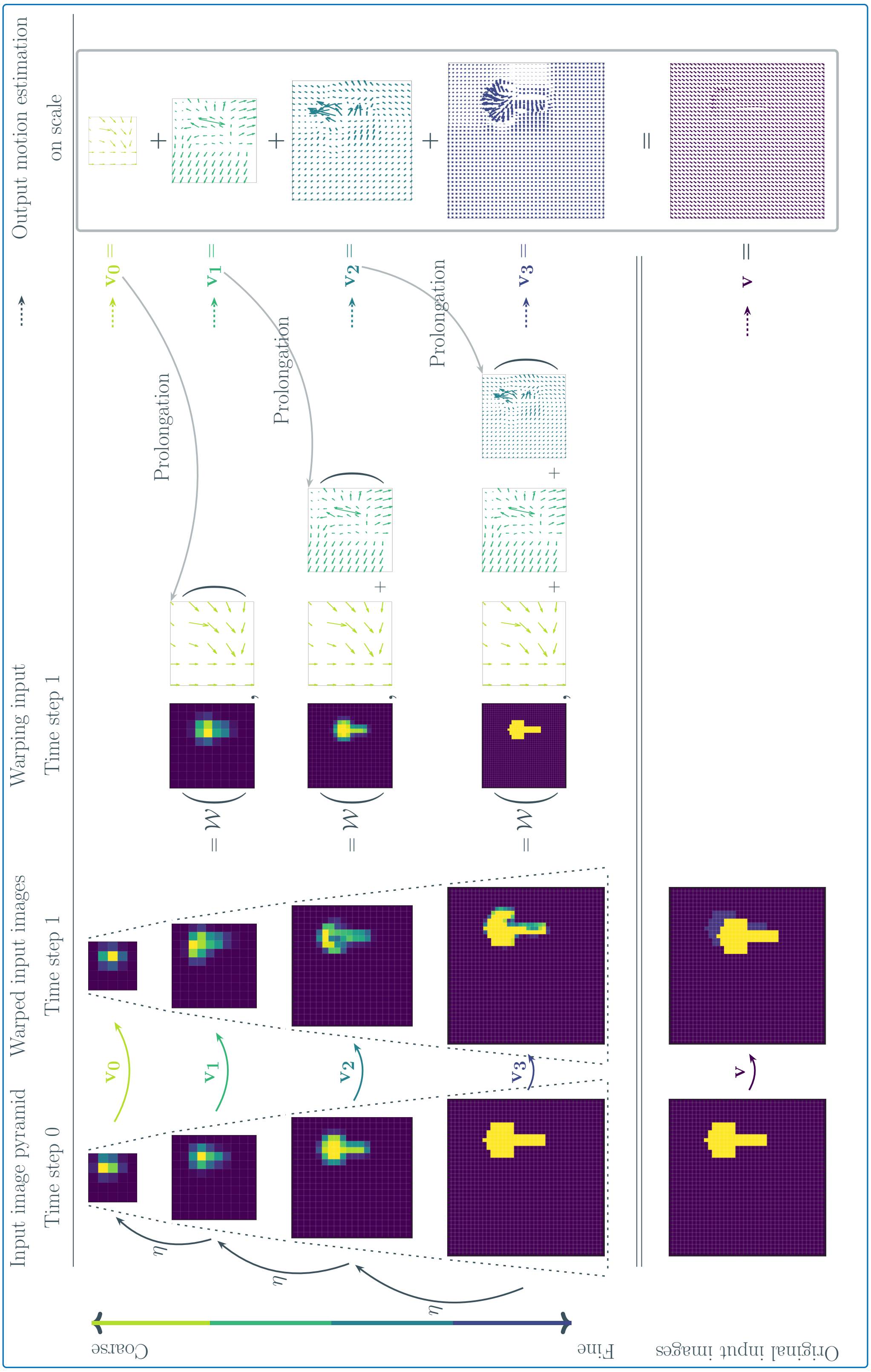


Figure 4.3.: Schematic diagram of the coarse-to-fine strategy with warping. First, image pyramids are derived for both time steps from the finest to the coarsest scale by applying the subsampling factor η. Afterwards, the motion estimation itself starts. Beginning on the coarsest scale, a motion estimate is computed. After prolongation of the estimate v_0 to the next finer scale, the result from the coarser scale is used for (reverse) warping of the input image at the corresponding scale in the image pyramid of the second time step. Next, a motion increment is computed between the input image at time step zero and the warped image at time step one. This process of prolongation, warping and increment computation is iterated over all scales. The output is defined as the sum of motion estimates over all scales.

such that

$$\operatorname{prox}_{\sigma \tilde{f}_2^*}(u) = \operatorname{prox}_{\sigma f_2^*}(u + \sigma C v_{\text{prev}}),$$

i.e., we have a slightly shifted argument making sure that smoothness is enforced throughout the added-up motion estimate and not only on one scale.

Multiscale for the mass conservation constraint

Remember that, in case of the MC constraint, we cannot use warping. First, the displacement fields might not be invertible if sources and sinks exist and moreover, the process of image warping itself assumes the pixels' brightness constancy. However, we need a multiscale strategy in order to deal with large displacements. We thus implement a coarse-to-fine strategy, where the motion estimate on each scale is initialized by the prolongated result of the last scale [111] and then updated after the increment is computed. The algorithm is given in Algorithm 5 and the process is illustrated in Figure 4.4.
We now briefly show that the implementation of PDHG for computation of increments is again straightforward. Consider the saddle point formulation on scale z

$$\min_{\delta_z} \max_{y_z} f_1(v_{z-1} + \delta_z) + \Big\langle y_z, C(v_{z-1} + \delta_z) \Big\rangle - f_2^*(y_z),$$

where v_{z-1} is known and fixed and δ_z is the searched-for increment. The steps for PDHG algorithm are then as follows (we omit the scale indices to reduce the notational complexity):

$$\begin{aligned} v + \delta^{k+1} &= \operatorname{prox}_{\tau f_1}\left(v + \delta^k - \tau C^* \tilde{y}^k\right), && (4.10)\\ y^{k+1} &= \operatorname{prox}_{\sigma f_2^*}\Big(y^k + \sigma C\left(v + \delta^{k+1}\right)\Big), && (4.11)\\ \tilde{y}^{k+1} &= y^{k+1} + \theta\left(y^{k+1} - y^k\right). \end{aligned}$$

The extrapolation step is not affected by the fact that we compute increments only. Also in the dual update step (4.11), the summand $\sigma C v$ is used in every iteration but has to be computed only once in advance. If C contains, e.g., a differential operator, the extended argument of the proximal mapping ensures smoothness of the whole iterate $v + \delta$ instead of the increment only. For the primal update step (4.10), we observe a correction by v after the proximal step with shifted argument. For the formulation (4.9), we have $f_1 = 0$ such that the proximal operator is the identity. In this case, v cancels out in (4.10).

4.3. Numerical validation on academic test cases

This section tests the motion estimation algorithms derived in the previous sections on some synthetic academic test cases that are inspired by medical applications. We choose phantoms that resemble images obtained by an MPI scanner, i.e., we use background-free images and compare the algorithms in terms of accuracy of obtained motion fields, run time and sensitivity to noise. Note that background-free images pose a special challenge in motion estimation, as the edge information is very limited.

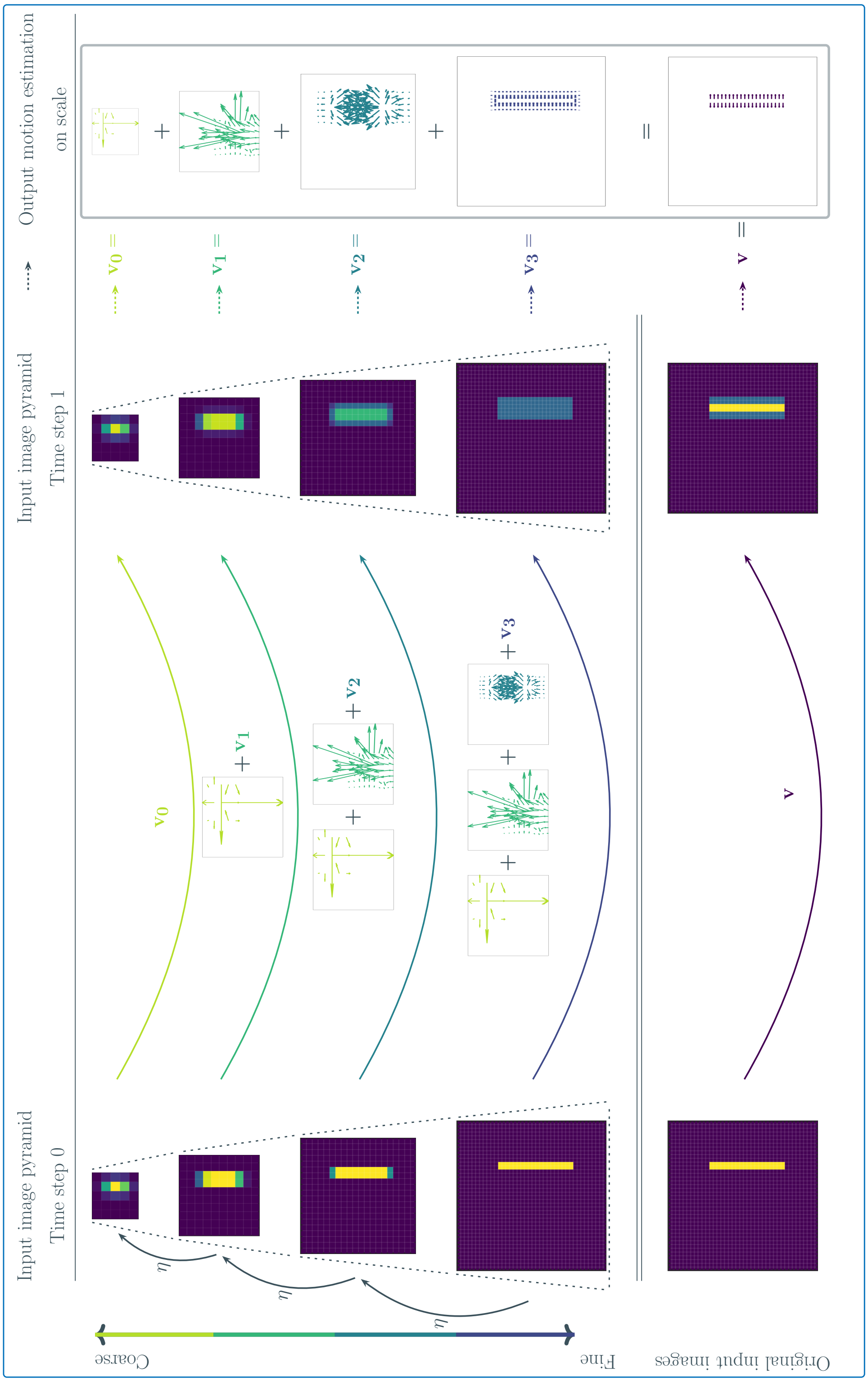


Figure 4.4.: Schematic diagram of the coarse-to-fine strategy without warping. First, image pyramids are derived for both time steps from the finest to the coarsest scale by applying the subsampling factor η. Afterwards, the motion estimation itself starts. Beginning on the coarsest scale, a motion estimate is computed. The motion estimate is prolongated to the next finer scale and used as fixed part of the motion estimate on that scale. Next, the motion increment is computed between the input image at time step zero and the image at time step one, given the fixed part of the motion from the previous scale. This process of prolongation and increment computation is iterated over all scales. The output is defined as the sum of motion estimates over all scales.

Algorithm 5 Motion estimation with multiscale

Input: image $c = c_Z$, initial motion field guess $v = v_Z$, number of scales Z, down-sampling factor η, $\tilde{v}_{-1} = 0$
for z=Z-1,...,0 **do**
  $c_z = \text{imresize}\,(\eta, c_{z+1})$ *% build image pyramid*
end for
for z=0,..,Z **do**
  $v_z = \text{imresize}\left(\eta^{Z-z}, v_Z\right) - \sum_{k<z} \text{imresize}\left(\eta^{k-z}, v_k\right)$ *% initial increments*
end for
for z=0,...,Z-1 **do**
  v_z = detect motion(c_z, initial value v_z, prior $\tilde{v}_{z-1}$)
  $\tilde{v}_z = \text{imresize}\,(\eta^{-1}, v_z + \tilde{v}_{z-1})$
end for
v_Z = detect motion(c_Z, initial value v_Z, prior $\tilde{v}_{Z-1}$)

More particularly, we first comment on quality measures for displacement fields in Section 4.3.1. In Section 4.3.2 we then compare the PDHG algorithm for OF constrained motion estimation (PDHG-OF) and the PDHG algorithm for MC constrained motion estimation (PDHG-MC) for small displacements. Afterwards, we consider larger displacements and compare the application of PDHG-OF and PDHG-MC directly on a single resolution scale to application within multiscale schemes. In the succeeding Section 4.3.3, we analyze the effect of gaussian smoothing in order to limit the influence of noise on the input data. Furthermore, we introduce a structural prior for motion estimation in Section 4.3.4. We conclude this section by analyzing the difficulty of parameter choices and the run time as well as computational costs of the algorithmic approach in Section 4.3.5.

4.3.1. Quality measures for displacement fields

We compare the flow fields obtained by the different algorithms by two means. First, we consider the euclidian distance between the ground truth vector field and the computed displacement field relative to the norm of the ground truth. This value gives an indication if the magnitude of the displacement is correctly computed. Second, we consider the Averaged Angular Error (AAE) between the ground truth and the computed displacement field. That is, we compute the angle between (normalized) ground truth and (normalized) computed motion in every voxel and then average over all values. This measure indicates the correctness of the direction of the computed displacements. Thus, the AAE also helps in interpreting the euclidian distance. If we observe a high euclidian distance in combination with a low AAE, the magnitude of motion is not accurately computed.
There are several difficulties with the quality assessment for displacement fields. First, we assume by the choice of our motion estimation algorithms, that the displacement between the image frames is linear as this is the only displacement we can compute. This assumption is reasonable as long as the time steps are sufficiently small.
The second and more important problem arises from the nature of the ground truth. Remember that we consider MPI images that typically have large empty background regions. In these areas, no information on the motion is available. As a consequence,

different displacement fields might describe the correct displacement and the exact output of motion estimation algorithms is mainly determined by the choice of regularization terms. Moreover, what we conceive as the best result might be highly subjective. The problem of different underlying motion fields is illustrated by Figure 4.5, where either of the three displacement fields might be the ground truth for the motion in the input image sequence.

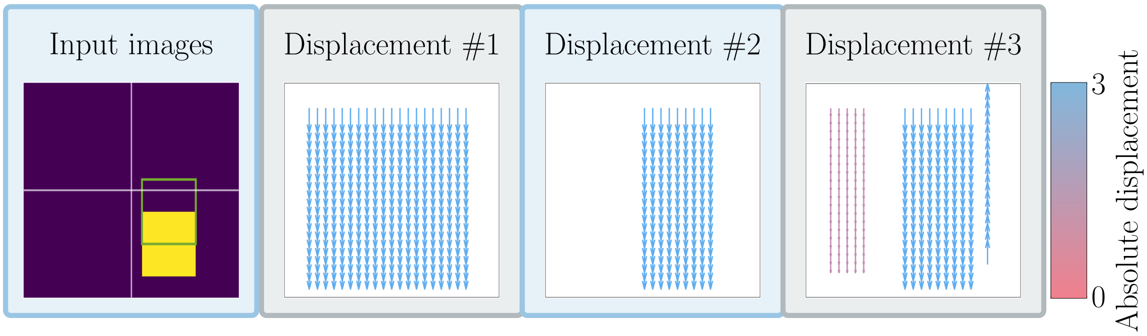


Figure 4.5.: Motion estimation without regularization does not result in unique solutions. Different displacement fields may correctly describe the displacement of the phantom in the input images sequence (left column). The phantom at the first time step is indicated by the green rectangle, the displacement between first and second time step might be caused by either of the three displacement fields.

By the choice of TV regularization, we accept severe fill-in artifacts in empty regions leading to spatially homogeneous motion estimates. Both AAE and L^2-norm error are computed on the full displacement field and not only in certain regions. This is due to the fact that defining a region-of-interest is a highly complex task if not impossible for some phantoms and again poses a subjective assumption on the correct displacement field. We have to keep in mind that error estimates on displacement fields might not accurately describe the quality of the fields.

4.3.2. Basic phantom - coarse-to-fine strategy

We start by considering a very simple phantom that consists of a square moving downwards with a velocity of one voxel per frame. We compute the displacement field by PDHG-OF and PDHG-MC. As can be seen in Figure 4.6, both algorithms yield good displacement fields even without applying a multiscale approach for this small displacement (top row). The relative euclidian distances between the displacement fields and the ground truth displacement are 0.08 and 0.05 for PDHG-OF and PDHG-MC, respectively. However, even for a slightly larger displacement of two voxels per frame instead of one, both algorithms can only accurately estimate the direction of the displacement, see Figure 4.6 (middle row), but not the absolute value. Both methods tend to underestimate the displacement, such that best achievable relative euclidian distances are now 0.42 and 0.09 for PDHG-OF and PDHG-MC. This means a five times higher error for PDHG-OF and a twice as high error for PDHG-MC.

By applying a multiscale and warping scheme (in case of PDHG-OF) or a multiscale scheme (in case of PDHG-MC), the euclidian distance can be reduced to $1.2 \cdot 10^{-4}$ and $1.4 \cdot 10^{-3}$, respectively, i.e., the displacement field is accurately estimated. This is also

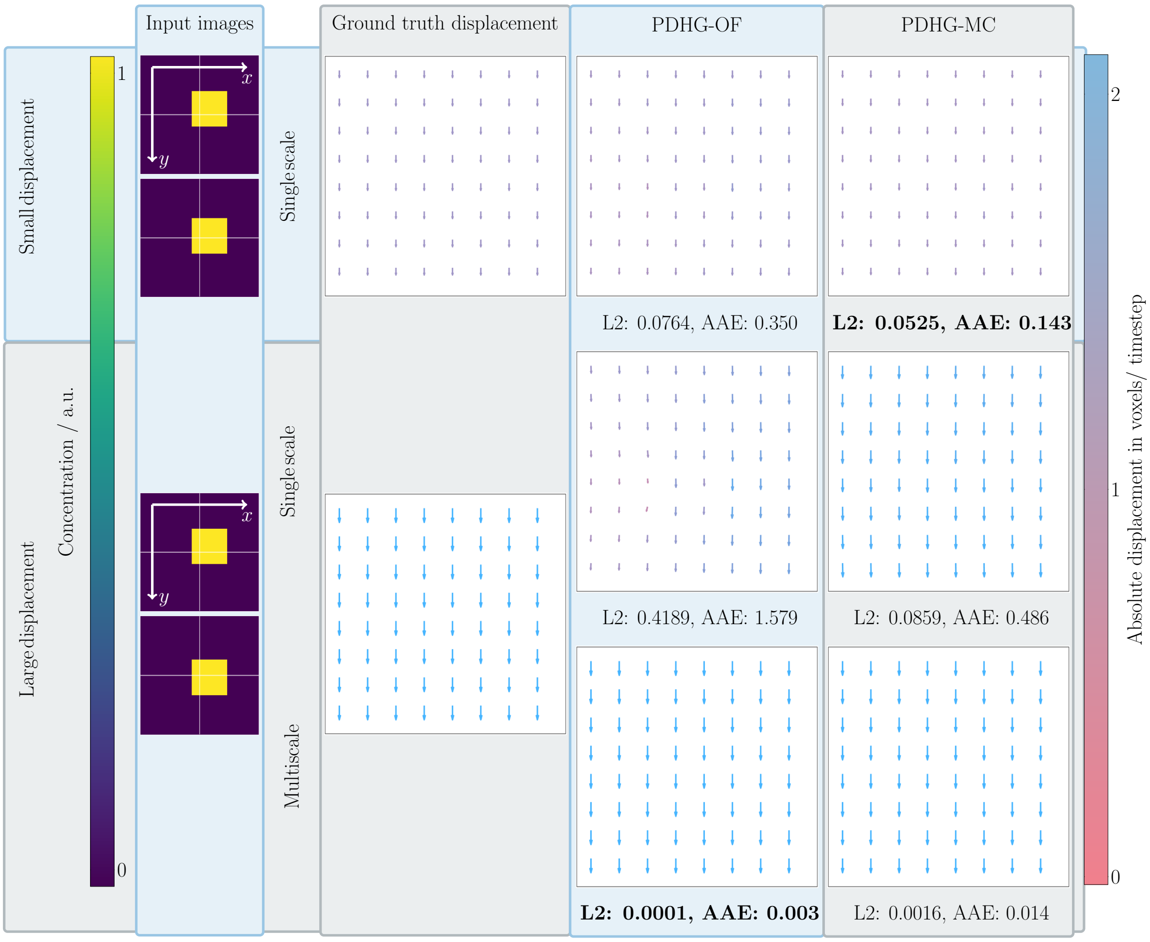
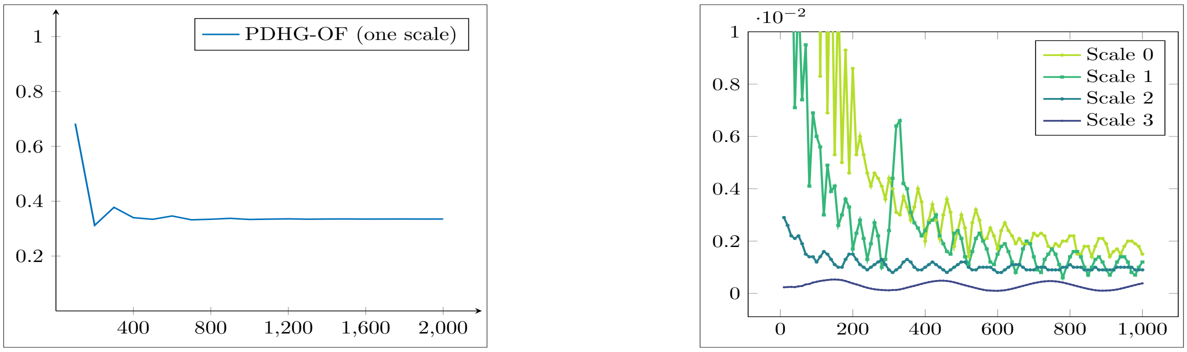


Figure 4.6.: Computed motion estimates for the basic phantom show the necessity of applying a multiscale scheme in combination with the primal-dual reconstruction algorithms. The input images of the basic phantom are shown in the left column. The ground truth displacement is one voxel in x-direction for the upper row (small displacement) and two voxels for the middle and bottom row (large displacement). Both algorithms, i.e., PDHG-OF and PDHG-MC, are able to recover the displacement field quite accurately without applying a multiscale approach for the small displacement (upper row). However, for the large displacement, the quality of the estimates decays without multiscale scheme applied (middle row). If a coarse-to-fine strategy is applied, the displacement fields have significantly lower L^2 and AAE error (bottom row).

confirmed by the results shown in Figure 4.6 (bottom row). We note that for the larger displacement of two voxels, we obtain results with significantly higher precision than PDHG without multiscale can obtain even for the small displacement.

The minimization of the objective value in a multiscale scheme compared to a single scale scheme is depicted in Figure 4.7. The minimum value reached by the algorithm using only one scale is significantly higher than the one reached by the multiscale scheme. We conclude that using a multiscale approach can also avoid being trapped in a local minimum (although the result obtained by the multiscale scheme might also be a local minimum only, however, it is a significantly better one).

After this brief proof of concept for using a multiscale strategy, we denote the multiscale and warping scheme using PDHG with the OF constraint on every scale by PDHG-OF and the multiscale approach using PDHG for the MC constraint on every scale by PDHG-MC.



(a) Objective value for reconstruction on the finest scale only.

(b) Objective value for reconstruction by a multiscale scheme.

Figure 4.7.: Decrease of the objective value for the basic phantom for a single scale scheme (a) and a multiscale scheme (b). When applying PDHG-OF on the finest scale only, the algorithm converges fast but the minimum objective value is at approx. 0.33. This is a local minimum, as we can see by applying a multiscale algorithm. The minimum objective value reached by the multiscale scheme is approx. 0.0005. Convergence on coarse scales needs more iterations, but the iterations on these scales are not costly. On finer scales, only few iterations are needed for the refinement and only small improvements in terms of the objective value are achieved.

4.3.3. Rotation phantom - gaussian smoothing

We now consider a phantom that meets the OF constraint. The phantom consists of a ball that moves along a spiral, i.e., we have movement along a circular path and a slow displacement along the z-direction. We consider four discrete time steps, projections onto the x-y-domain are shown in Figure 4.8 (left column). The absolute displacements in voxels are $3/2/1$ in x-direction and $-1/-2/-3$ in y-direction and thus exceed one voxel per frame.

Applying the coarse-to-fine algorithms to noise-free input images yields good results, as is shown in Figure 4.8. The direction as well as magnitude of the displacements is well recovered. For PDHG-OF, the AAE is below one degree, for PDHG-MC it is below six degrees, as can be seen in Table 4.1. We notice that PDHG-OF performs better than PDHG-MC. For the basic phantom, the opposite was true. Having in mind medical imaging applications, in a realistic setting noise on the input images will be present. We consider four different noise levels.

For noise level one, we add gaussian noise with standard deviation of 2.5% of the maximum intensity in the images, for noise level two we consider a standard deviation of 5%, for noise level three 10% and for noise level four, we have a standard deviation of 20%.

Let us first consider PDHG-OF. For noise level one, the results seem similar to the noise-free case visually. Although errors in angle and magnitude are approximately three times higher (see the errors in Table 4.1), the true displacement is almost recovered. For noise level two, we now also observe larger angular inaccuracies, cf. Figure 4.9. This visual impression is confirmed by the AAE and L^2-discrepancy between the estimated flow fields and the ground truth in Table 4.1. For noise level three, PDHG-OF still yields reasonable estimates, whereas the motion estimates for images at noise level four are not resembling the ground truth anymore.

Considering PDHG-MC, the algorithm faces severe problems under noise. While the

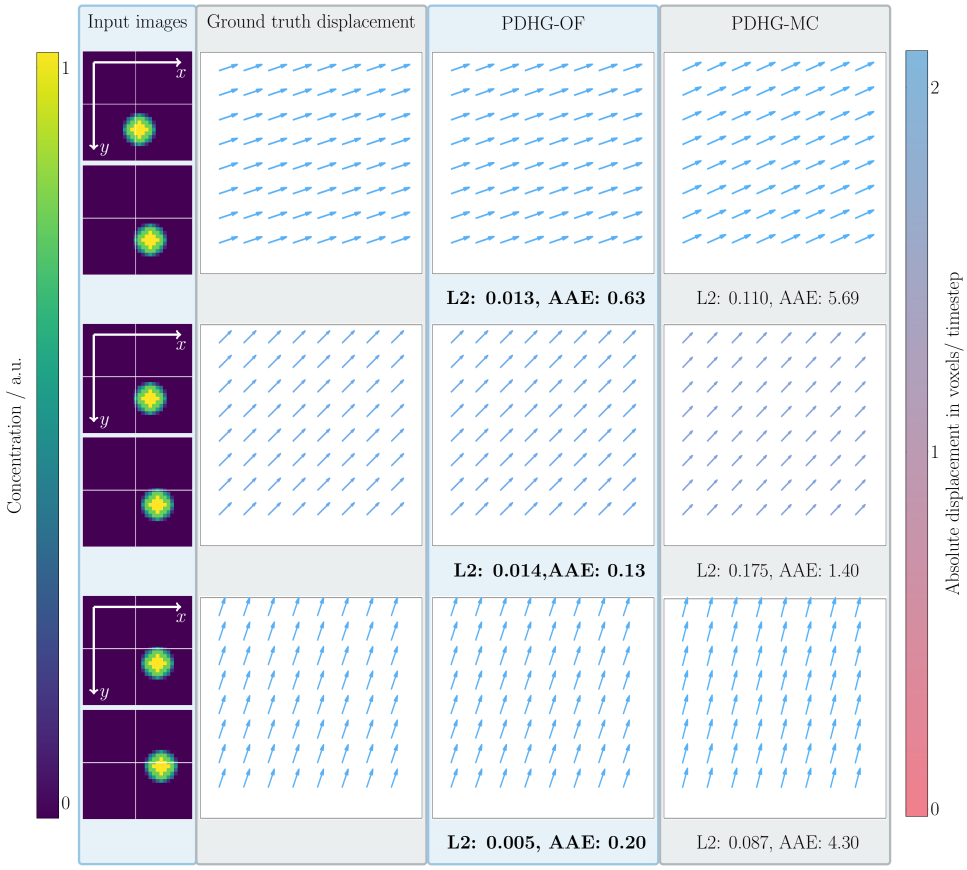


Figure 4.8.: Computed motion estimates for the rotation phantom show significantly lower angular error for PDHG-OF compared to PDHG-MC. The rotation phantom (noise-free) is displayed as projections onto the x-y-domain in the left column. The second column depicts the ground truth displacement fields whereas the flow fields in columns three and four are computed by PDHG-OF and PDHG-MC, respectively. Both algorithms yield results with comparable L^2 error, but PDHG-OF has significantly lower angular error.

direction of the flow is still perceptible (but already has an angular error of more than ten degrees), the magnitude of the computed displacement fields does not match the ground truth already for noise level one. The quality of the resulting flow fields is comparable for noise levels one to three and degrades more for noise level four.

The resulting displacement fields for levels two and three are depicted in Figure 4.9 and Figure 4.10, respectively.

We emphasize that the visually perceived noise level in the input images is low compared to standard medical images. For all noise levels, the main phantom can be clearly distinguished from the background. Up to a certain (algorithm dependent) point, the error of the computed displacement field grows almost linearly with the noise on the input image. However, at a certain noise level the algorithm is not able to produce meaningful results. Having almost noise-free background significantly improves the results of the motion estimation task. Thus, we expect motion estimates on MPI image sequences reconstructed by primal-dual methods with FL regularization to achieve significantly higher quality compared to motion estimates based on image sequences reconstructed by standard Kaczmarz method (cf. Section 3.3 for a comparison of different MPI reconstruction techniques).

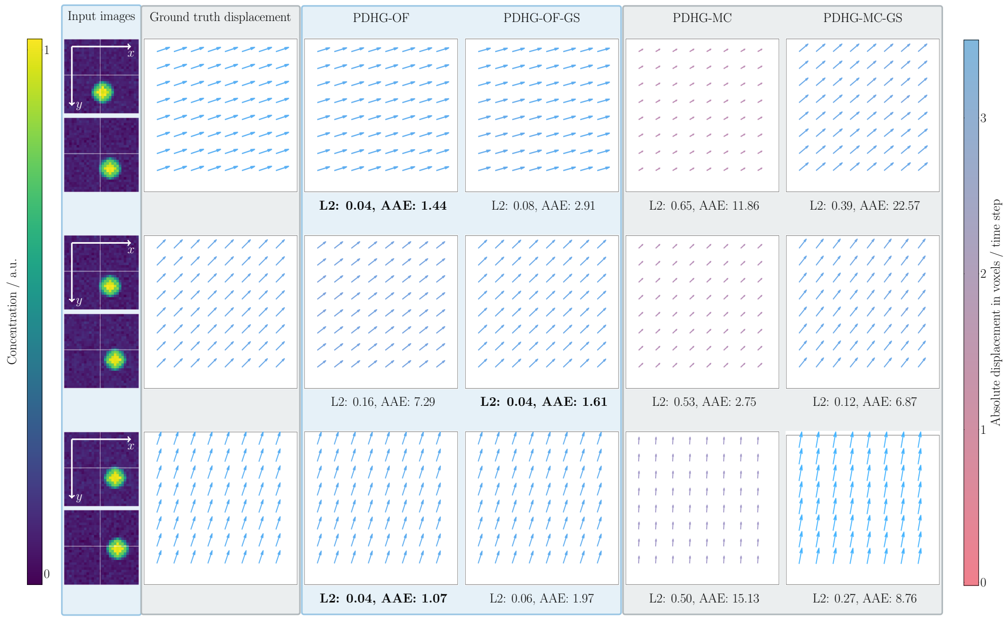


Figure 4.9.: The displacement fields for the rotation phantom under noise are significantly better when using PDHG-OF instead of PDHG-MC for the 5% noise level. The phantom (noise level two, i.e., 5% noise) is displayed as projections onto the x-y-domain in the left column. PDHG-OF (column three) recovers the true displacement (ground truth in second column) well, the impact of the gaussian smoothing (column four) is minor. The estimates by PDHG-MC (column five) are significantly worse with respect to L^2-error and AAE. Without applying gaussian smoothing, the magnitude of motion is not recovered at all. The angular errors are high compared to the ones by PDHG-OF and even higher when applying gaussian smoothing (column six).

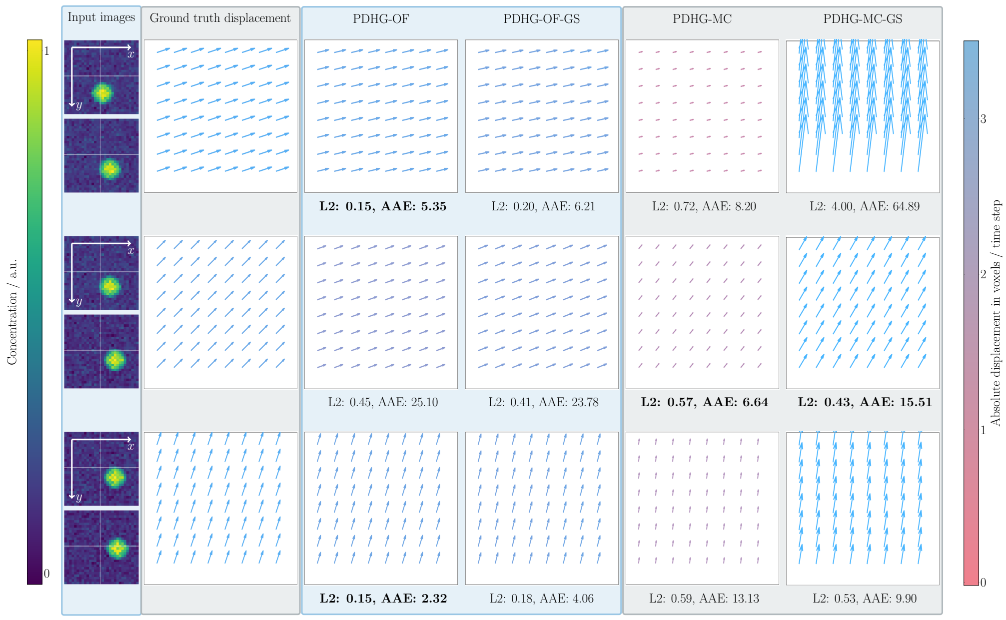


Figure 4.10.: The displacement fields for the rotation phantom under noise suffer from severe artifacts for the 10% noise level. The phantom (noise level three, i.e., 10% noise) is displayed as projections onto the x-y-domain in the left column. PDHG-OF (column three) recovers the true displacement (ground truth in second column) better than the other schemes for time steps one and three, the impact of the gaussian smoothing (column four) is minor. For time step two, the estimates by PDHG-OF and PDHG-OF-GS are worse than the results by PDHG-MC (column five) and PDHG-MC-GS (column six). PDHG-MC underestimates the magnitude of motion, whereas PDHG-MC-GS overestimates it.

Table 4.1.: The mean AAE and the relative euclidian distance between estimated displacement fields and the ground truth displacement for the synthetic rotation phantom (averaged over three time steps, i.e., three different displacements, and 100 runs of each algorithm) for different noise levels and algorithms.

	Noise-free 0%		Level 1 2.5%		Level 2 5%		Level 3 10%		Level 4 20%	
Algorithm	L^2	AAE	L^2	AAE	L^2	AAE	L^2	AAE	L^2	AAE
PDHG-OF	**0.01**	0.32	**0.03**	**0.66**	0.09	2.43	**0.24**	8.52	0.73	41.34
PDHG-OF-GS	**0.01**	**0.28**	**0.04**	0.85	**0.06**	**1.57**	**0.23**	**7.63**	**0.68**	36.52
PDHG-MC	0.13	3.80	0.56	10.93	0.60	12.03	0.67	13.49	0.79	**19.44**
PDHG-MC-GS	0.14	4.74	0.42	10.65	0.65	15.77	1.27	24.34	0.76	**19.55**

Moreover, the quality of the resulting motion estimates varies strongly for different representations of noise. The AAE and L^2-error values in Table 4.1 are mean values over 100 runs and have high standard deviation. The standard deviations relative to the mean errors are presented in Table 4.2.

In order to obtain more stable results, we apply gaussian smoothing as proposed for the CLG method [33] (see Section 4.2). For PDHG-OF the implementation is straightforward. We apply gaussian smoothing with a kernel with standard deviation 0.5 to the spatial image derivatives and denote the algorithm by PDHG-OF-GS. For PDHG-MC, the image derivatives are not used directly in the computational scheme. Due to the form of the MC constraint, the derivative of cv is used within the algorithm. Smoothing the differential of the product allows undesired deviations of v. Instead, we apply gaussian smoothing already to the input images c. This has the advantage of not smoothing v within the algorithm but by smoothing the image instead of the derivatives, we do not only cancel noise but also smooth edges and thus impede motion estimation.

PDHG-OF-GS improves the motion estimation under noise compared to PDHG-OF. The AAE is significantly lower compared to the one achieved by PDHG-OF for noise levels two to four (see Table 4.1). As expected, the differences are minor for the noise-free case as well as for noise level one.

PDHG-MC has even worse results if combined with gaussian smoothing, see PDHG-MC-GS in Table 4.1 and Table 4.2. As explained above, we relate this to smoothing at the wrong end. Resulting displacement fields for levels two and three are depicted again in Figure 4.9 and Figure 4.10.

We conclude that gaussian smoothing is useful when applying PDHG-OF to potentially noisy data, but cannot be applied when using PDHG-MC.

Table 4.2.: The mean L^2-error and AAE as well as their standard deviations relative to the mean value between estimated displacement fields and the ground truth displacement for the synthetic rotation phantom for different noise levels and algorithms.

	Level 1 2.5%				Level 2 5%			
Algorithm	L^2	σ	AAE	σ	L^2	σ	AAE	σ
PDHG-OF	**0.03**	21%	**0.66**	70%	0.09	41%	2.43	72%
PDHG-OF-GS	**0.04**	44%	0.85	72%	**0.06**	26%	**1.57**	60%
PDHG-MC	0.56	**2%**	10.93	**27%**	0.60	**3%**	12.03	**37%**
PDHG-MC-GS	0.42	65%	10.65	62%	0.65	111%	15.77	46%

The experiments on the rotation phantom show the importance of having clean input images not corrupted by noise. Therefore, when considering MPI image sequences, we choose a regularization method that cancels noise effectively. Moreover, it is important to preserve edges in the input images. Trading noise for edges was done by using gaussian smoothing and did not yield the desired outcome. In the next section, we propose a regularization method incorporating edge information explicitly. That method is derived for the PDHG-MC scheme, where the gaussian smoothing is not applicable.

4.3.4. A structural prior for motion estimation

MPI data is often combined with MRI images to obtain background information on the considered tissue in order to enhance interpretability of images and to observe functionality of organs. Those images can be laid underneath the MPI images in a last processing step, but we can also use this prior knowledge to improve image reconstruction and motion estimation. In [15], it was shown that a structural prior based on information provided by MRI significantly improves the image reconstruction in MPI. A structural prior based on MRI images for PET reconstruction was proposed in [53]. Therein, the authors propose a convex prior, which reduces locally to TV if no a-priori knowledge is available. In this work, we consider a structural prior not for the image reconstruction task but for the motion estimation task in order to obtain a more robust algorithm with respect to noisy input data. The remainder of the section is structured as follows. We first derive the prior theoretically and comment on its integration into the existing PDHG scheme for motion estimation. We then present a brief numerical proof of concept based on academic examples.

Let $B : \Omega \to \mathbb{R}$ describe a background image obtained by, e. g., an MRI scan. The edges of such a background image can then be described by the gradient of the image, in particular the gradient ∇B is orthogonal to the edges of the background image B. Let us imagine the application of blood flow imaging and a background image which depicts the patients vascular system. Physically, a blood flow is likely to occur in direction of the blood vessels, i.e., aligned with the vessels' walls. A flow orthogonal to the walls or through the walls of the vessels is unlikely. More precisely, a flow of small magnitude might occur in that direction, however, a strong flow in that direction is very unlikely. Let us now incorporate this idea into a penalty term. We propose a penalty term $\mathcal{P} : \mathrm{BV}(\Omega)^n \to \mathbb{R}$ given by

$$\mathcal{P}\big(v(\cdot,t)\big) = P\big(v(\cdot,t)\big)\,\|v(\cdot,t)\|\,,$$

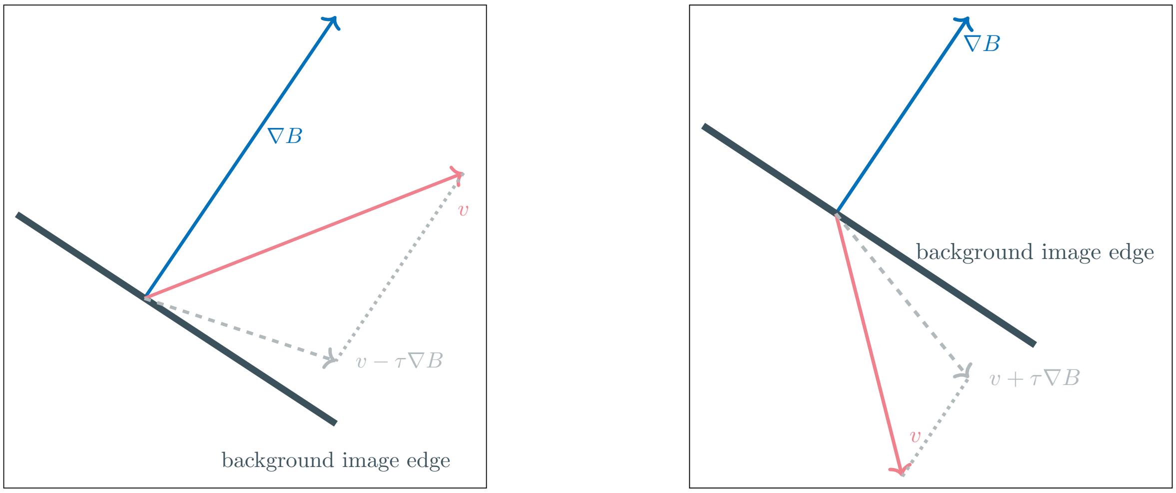
where $P : \mathrm{BV}(\Omega)^n \to \mathbb{R}$ accounts for the directional penalization. The magnitude of the flow is taken into account by the multiplication with the norm of the displacement v. We now define the term P based on the following considerations. A flow v, which is orthogonal to ∇B is parallel to the edges of the background image and thus parallel to, e.g., walls of blood vessels. This flow is highly plausible and should not be penalized. In contrast, a flow v, which is parallel to the gradient of B is orthogonal to edges of the background image. This is the unlikely case of a flow through an edge, which marks, e.g., the vessel walls. We thus want to penalize such a flow.

We define
$$P\big(v(x,t)\big) = \big|\cos\big(\nabla B(x), v(x,t)\big)\big|\,,$$
and obtain
$$\big|\cos\big(\nabla B(x), v(x,t)\big)\big| = \left|\left\langle \frac{\nabla B(x)}{\|\nabla B(x)\|}, \frac{v(x,t)}{\|v(x,t)\|}\right\rangle\right| = \frac{1}{\|v(x,t)\|}\left|\left\langle \frac{\nabla B(x)}{\|\nabla B(x)\|}, v(x,t)\right\rangle\right|,$$
which yields
$$\mathcal{P}\big(v(x,t)\big) = \left|\left\langle \frac{\nabla B(x)}{\|\nabla B(x)\|}, v(x,t)\right\rangle\right|.$$
This penalty term is proper, convex and lower semicontinuous. Moreover, it reduces locally to zero if no background information is available, i.e., if no edges are present in the background image. As we combine it with a TV prior in our setting, the overall penalty term then reduces to TV locally. It can be easily integrated into the PDHG scheme for motion estimation with the MC constraint as follows. In this setting, we note that (using the same notation as in Section 4.2)
$$f_1(v) = \int_0^T \left|\left\langle \frac{\nabla B}{\|\nabla B\|}, v\,(\cdot,t)\right\rangle\right| \mathrm{d}t.$$
Simple calculations yield the proximal operator
$$\mathrm{prox}_{\tau f_1}(u) = \begin{cases} u + \tau\frac{\nabla B}{\|\nabla B\|}, & \text{if } \left\langle \frac{\nabla B}{\|\nabla B\|}, u\right\rangle < -\tau \\ u, & \text{if } -\tau \leq \left\langle \frac{\nabla B}{\|\nabla B\|}, u\right\rangle \leq \tau \\ u - \tau\frac{\nabla B}{\|\nabla B\|}, & \text{if } \left\langle \frac{\nabla B}{\|\nabla B\|}, u\right\rangle > \tau \end{cases}$$
defined pointwise in time. For illustration of the operator, let us consider the following example. Assuming u to be normalized, the first case, $\left\langle \frac{\nabla B}{\|\nabla B\|}, u\right\rangle < -\tau$ corresponds to an obtuse angle between the normalized gradient of the background image and the motion estimate u. In this case, shifting the motion estimate into the direction of $-\nabla B$ leads to a more acute angle and thus aligns the direction of motion with the edges in the background image. The opposite holds for the third case. This behavior is illustrated by Figure 4.11. The parameter τ defines a region where we accept the angle between motion estimate and image gradient.

Unfortunately, it is not straightforward to integrate a similar penalty into the PDHG-OF scheme. The prior cannot be integrated into the primal part of the functional, as the resulting functional is then not prox-tractable, i.e., the proximal operator is not easy to compute. Thus, we need to consider another algorithmic scheme. As this short section serves as a first proof of concept for integration of a structural prior into a motion estimation scheme, we decide to limit ourselves to the PDHG-MC scheme and leave the generalization to other motion models for future work.

Numerical evaluation

To evaluate the angular penalization based on background images, we first consider a simple toy example. The phantom consists of two time frames. In each frame, a cuboid



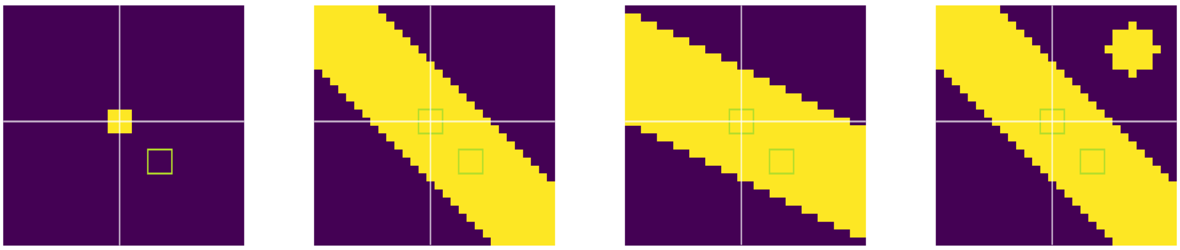
(a) Acute angle: $\left\langle \frac{\nabla B}{\|\nabla B\|}, \frac{v}{\|v\|} \right\rangle > 0.$

(b) Obtuse angle: $\left\langle \frac{\nabla B}{\|\nabla B\|}, \frac{v}{\|v\|} \right\rangle < 0.$

Figure 4.11.: The proximal operator of the directional penalty term for motion estimation for acute and obtuse angles. For acute angles in (a), the proximal operator enlarges the angle between the background gradient ∇B (blue) and the motion estimate v (red). For obtuse angles in (b) the angle is reduced for τ sufficiently small.

based on a square of size 3×3 voxels is depicted. The square is placed randomly on a domain of 21×21 voxels within the image domain of 30×30 voxels for the first time step. For the second time step, it is placed randomly within a 9×9 neighborhood of the center of the first frames' position. This yields 32 different possible directions of the displacement with a magnitude varying from 0 to 5.66, in total we have 81 different displacements. In the following, we compute motion estimates for 80 different displacements (leaving out the case of no displacement) for three different background versions and compare the effect. When computing motion estimates without any structural prior, PDHG-MC yields a mean AAE of 8.5 degrees. Note that we have a noise-free setting for this first experiment.

First, we use optimal background information for the searched-for displacement. More particularly, the background images consist of a path connecting the two squares. An illustrative background image for a specific phantom is depicted in Figure 4.12b.



(a) An exemplary phantom.

(b) The optimal background image.

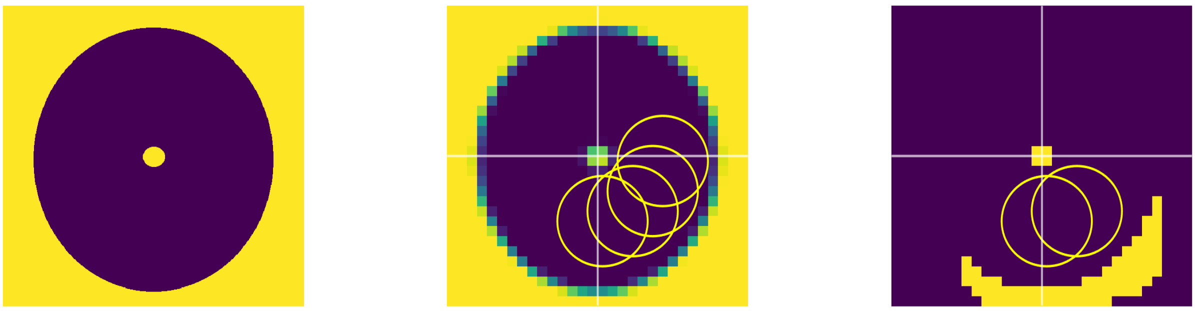
(c) Background with slightly altered slope.

(d) Background with distraction.

Figure 4.12.: An exemplary phantom and background options for evaluation of the structural penalty term for motion estimation based on a background image and its edges.

In this setting, the mean AAE over the 80 different ground truth displacement mappings (leaving out the case of no displacement) is reduced from 8.5 to 4.37, i.e., by 49% by incorporating the background information. The L^2-norm error thereby stays constant. Next, we slightly change the background information by using a different slope of the tunnel in which the phantom squares move, see Figure 4.12c. The slope is altered by increasing or decreasing it randomly by 1. In this case, the mean AAE can be reduced to 7.15, which translates to a reduction by 16%. Further, we add two balls as distraction elements to the background, where one of them might coincide with the tunnel, see Figure 4.12d. In this setting, the AAE can be reduced to 5.14, i.e., by 40%. This experiment serves as a proof of concept and shows the applicability of the penalty term. We observe that our structural prior works as expected, edges in the background image that are parallel to the flow significantly increase the angular accuracy of the flow estimates. However, also non-optimal background information in terms of edge direction compared to the flow direction yield an improvement.

Next, we consider a slightly more complex phantom, i.e., the rotation phantom from Section 4.3.3. We use this phantom in order to analyze the effect of a more complex background image as well as the effect of noise on the input images. The ball phantom moves along a circular path which we use as background information. The background image is depicted in Figure 4.13a. We observe edges in all different directions in the



(a) Full background, high resolution.

(b) Full background, reconstruction resolution.

(c) Framewise background.

Figure 4.13.: A background image for the rotation phantom. The full fine resolution background from (a) is downsampled to match the resolution of the phantom in (b). The position of the phantom during the first four frames is indicated by the circles. A partial background based on the position of the phantom during the first two frames (indicated by the circles) is depicted in (c).

background image depending on the spatial position of the edge. This poses a problem in combination with the TV prior which assumes a homogeneous direction of the motion estimates through space. The two priors thus pull the algorithmic solution to opposed states, leading to a solution where the direction of the motion estimates is influenced strongly by edges being far apart from the considered phantom position. In a noise-free scenario, application of the full background image as in Figure 4.13a does not lead to a significant change of the AAE if the parameter is small, for larger parameters the AAE even increases. In the low noise scenario, i.e., when considering 2.5% noise on the input images, the AAE even increases by a factor 7 from 10.93 to 75.50. Similar increases are observed for higher noise levels.

One possibility to deal with the background versus TV prior problem is to consider framewise background images instead of the full background. In this case, we define the region-of-interest as the neighborhood of the phantoms position in a certain frame and its succeeding frame. This neighborhood is then extended by gaussian smoothing to a slightly larger area. The original background image is only considered in this region and set to zero everywhere else. An exemplary framewise background image is depicted in Figure 4.13c. In this setting, we can still assume the motion fields to be spatially homogeneous without posing a contradiction to the background prior. In a noise-free scenario, the influence of the background penalty term is again negligible for small parameters. However, the mean AAE can be reduced from 10.93 to 7.75, i.e., by 29% in the low noise scenario (2.5%). With 5% noise on the input images the mean AAE decreases from 12.03 to 7.9, i.e., again by more than 30%. In the high noise scenario, i.e., with 10% noise added on the input images, the AAE with framewise background priors applied is only 8.6 instead of 13.49 previously. The decreases in terms of the AAE are summarized in Table 4.3. Consulting Table 4.1 we see that PDHG-MC now achieves results of similar quality compared to PDHG-OF for the higher noise levels.

Table 4.3.: Decrease of the mean AAE when incorporating framewise background images into the reconstruction scheme for the synthetic rotation phantom for different noise levels.

Noise level	Noise-free (0%)	Level 1 (2.5%)	Level 2 (5%)	Level 3 (10%)
AAE Reduction	0%	29.1%	34.3%	36.2%

Another possibility to deal with the oppositeness of the TV prior and the structural prior is to abandon the idea of spatially homogeneous flow fields and use another prior. A reasonable option would be a classical L^2-Tikhonov term on the motion fields. With this choice, we allow for spatially inhomogeneous motion fields. Practically, a drawback is the more difficult qualitative evaluation of the resulting fields as no ground truth is available in this case. We thus have to visually inspect the motion fields in general and can combine this with a local AAE criterion within the region-of-interest, e.g., the spatial position of the phantom. For our example above the background image consists of large homogeneous regions. In these regions, we do not have edges to incorporate into our penalty term. If combined with regularization of the gradient of the displacement fields, i.e., TV regularization, this is not restrictive as the fill-in effect transports knowledge about edge directions through the spatial region. If, however, combined with L^2-Tikhonov regularization, we do not have such a transport property. Thus, in each spatial position, only local edge information from that position exactly is available. Applying this idea to the synthetic data experiment above does not yield good results. The resulting flow fields have a local AAE of approximately 40 degrees both with and without background information incorporated. However, this might be due to the very sparse input images and background image in this example.

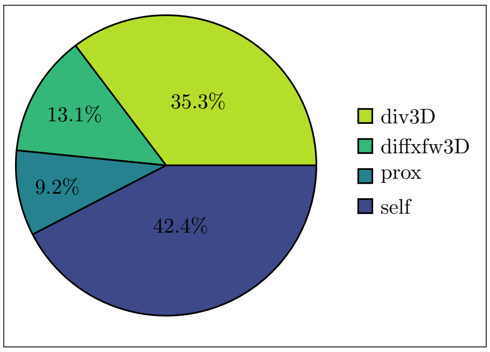
In summary, we note that incorporating background information for angular penalization of flow fields can increase the directional accuracy. If the temporal resolution is high enough such that we expect piecewise linear displacements and if we expect only one relevant direction of displacement per frame, we suggest using a framewise background image and a TV prior. This stabilizes the resulting flow fields in case of noisy input data.

If we expect a more complex flow, we should not use a TV prior. In this case, we can incorporate a full background image.

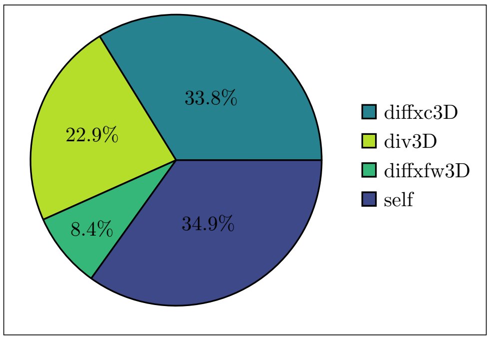
4.3.5. Parameter tuning, run time and cost

In this section, we briefly comment on the applicability of the motion estimation approach in practice, focusing on run time and computational bottlenecks.
However, we first consider the parameter tuning. Parameter tuning is a major challenge for variational optimization problems in general. A stable parameter choice is important for the applicability in practice, as the costs of permanently tuning a parameter are high and parameter tuning is a difficult task if no ground truth solution is available. Our algorithmic scheme has two main parameters, in particular we have the smoothness parameter β and the data fidelity parameter γ. In standard motion estimation schemes the parameter γ is set to one and fixed. In our case, we need γ explicitly when using the motion estimation scheme inside the joint optimization scheme and therefore, we tune both parameters jointly already in this section.
For the results in the previous sections, a parameter search on a grid with factor 10 was performed. The algorithm PDHG-OF has similar (difference up to factor 10) parameters for each noise level. For PDHG-MC, the optimal parameters differ by a factor up to 10^3, but only by a factor of 10 if gaussian smoothing is applied. For experiments on measured data, no ground truth motion field will be available in order to find the best parameters. Thus, we have to visually inspect each result in order to define the best one. A relatively stable parameter choice thus enables transfer of good parameters for one measurement to other measurements by the same scanner or in a similar environment.

Finally, we briefly analyze the computational costs related to the different motion estimation algorithms. Note that we did not focus on efficiency of the implemented algorithms yet. The main cost factor when applying PDHG algorithms for motion estimation are the differential operators used as is illustrated in Figure 4.14. They account for 48.4 % and 65.1% of the total cost of PDHG-OF and PDHG-MC, respectively. By porting the motion estimation task from MATLAB to C ++ the author of [50] was able to produce a speed-up of factor approximately 50. We thus expect potential for huge savings in this area.
Considering now the multiscale and warping and multiscale only scheme with 10^3 iterations on each scale, the costs on coarser scales are significantly lower than on finer scales. Figure 4.15 illustrates that more than 80% of computation time are spent on the finest scales for both schemes. The overall run time is approximately 30 s for both algorithmic solutions. Ideally, we implement a low-cost stopping criterion enabling early stopping on finer scales. In order to investigate this option, we consider the decay of the objective value for both schemes.
Figure 4.16 illustrates typical behavior of the objective value during the iterations on different scales for PDHG-OF and PDHG-MC. The values are scaled to the unit interval. The solutions on the different scales might jump between different solutions as illustrated by Figure 4.16a such that many stopping criteria will not be fulfilled. This jumping might occur for the OF constrained algorithm as well as for the MC constrained algorithm. In

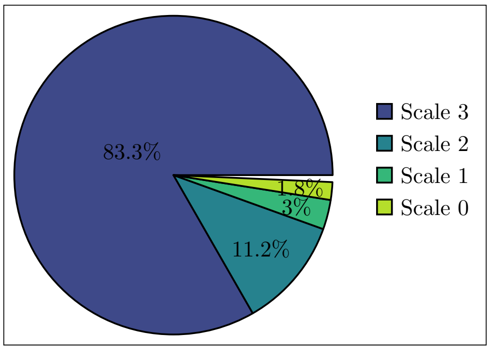


(a) PDHG-OF.

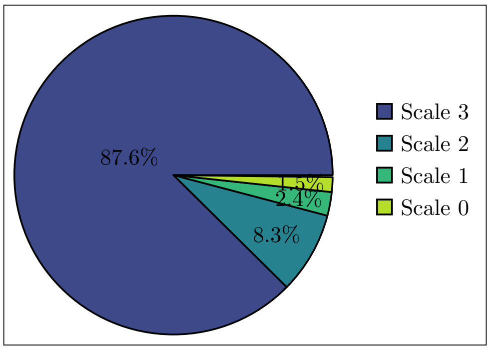


(b) PDHG-MC.

Figure 4.14.: Main sources of computational costs for motion estimation PDHG algorithms. The OF algorithm computes a spatial forward derivative (diffxfw3D) and a divergence (div3D), each of those functions is called 12.000 times and together they use 48.4% of the computation time. The proximal operator corresponding to the data term is also an important cost factor with approximately 10% of computational cost. The MC algorithm calls three differential functions, which in total make 65.1% of the run time.



(a) Multiscale and warping.

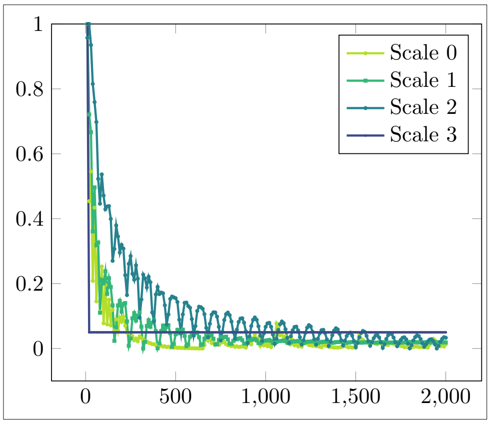


(b) Multiscale.

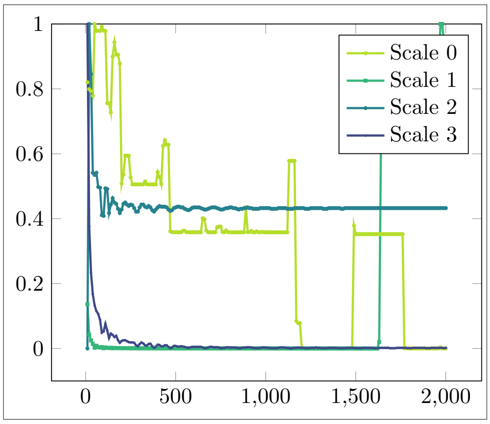
Figure 4.15.: Allocation of computation time to the different scales in multiscale motion estimation algorithms when applying a fixed number of iterations on each scale. More than 80% of run time are spent on computations on the finest scales, whereas computation on the coarsest scales take less than 2% of the overall cost.

order to obtain smoother convergence, we have to use smaller step sizes, which then again result in significantly more iterations needed. Moreover, they might trap us in worse local minima. However, depending on the exact problem to solve, a smooth decay of the objective value as in Figure 4.16b is also possible.

We test a stopping criterion based on evaluations of the objective value. For PDHG-MC, we stop the iterations on a certain scale, if either the maximum number of iterations is reached or if the standard deviation of the last 15 evaluations of the objective value is smaller than 0.01% of the mean objective value. For an illustrative run, this reduces the number of iterations needed on each scale from coarse to fine for PDHG-MC to 10^3, 690, 750 and 310 instead of previously 10^3 on each scale, respectively. However, the overall cost is only reduced by 16.7% as evaluations of the objective value are again quite costly. The allocation of run time to the different scales is more uniform, now only 61% are allotted to the finest scale (instead of 88%) and 28% are allotted to the second finest scale instead of previously only 8%.



(a) PDHG-OF.



(b) PDHG-MC.

Figure 4.16.: Exemplary behavior of the objective value during the iterations on the different scales on multiscale (and warping) schemes. The values are scaled to the unit interval.

For PDHG-OF, the standard deviation of the objective value does not reach that bound of 0.01%, as can also be guessed from the stronger jumping behavior in Figure 4.16a. Instead, we set the bound to 0.7% allowing for similar results compared to the setting with 10^3 iterations fixed per scale. The run time savings and allocation of run time change in a similar manner than above. Again, the run time saving is not as high as the evaluation of the objective value is costly.

Moreover, we frequently observe a different behavior: less iterations are used on coarser scales and thus more iterations are needed on finer scales. Fitting an optimal stopping criterion is beyond the scope of this work and should be done when a more efficient implementation can be used as basis.

4.4. Summary and discussion

This chapter focused on the problem of motion estimation from MPI image sequences. After deriving two different motion models, namely OF and MC, we derived a primal-dual splitting scheme for each motion model. Our main contribution in this chapter is contained in Section 4.2.3, where we derived a multiscale and warping scheme for the OF motion model and a multiscale scheme for the MC motion model. Both schemes include the aforementioned primal-dual algorithm for computation of increments on each scale. We extensively validated the algorithmic approaches and minor extensions thereof on synthetic data. The PDHG-OF yields better results with respect to the L^2-error and the AAE on our examples. However, the quality varies strongly depending on the exact input images and the direction and magnitude of the ground truth displacement. Both schemes reveal weaknesses under noise which cannot be overcome by gaussian smoothing. More precisely, we observe that PDHG-OF is slightly more robust with respect to noise on the input images, if gaussian smoothing is applied. For PDHG-MC it is not possible to integrate the gaussian smoothing efficiently. We thus proposed a structural prior using background information that could be obtained, e.g., by an MRI scan. This prior was incorporated into the PDHG-MC scheme and significantly improved the AAE

on noisy synthetic data. We observed difficulties when combining the structural prior with a TV prior for complex background images. Therefore, we proposed to use a region-of-interest background image for each frame, which results in enhanced angular accuracy. Finally, we investigated the computational costs of the algorithmic schemes. We found that differential operators constitute the main cost. Furthermore, we observed that the computational burden on the finest scales within the multiscale schemes are extremely high and a stopping criterion limiting those iterations to the bare minimum is needed.
Different directions for future research are of interest with respect to the motion estimation task. Incorporating prior knowledge from background information also for more complex flows, in combination with different priors and in case of the OF motion model is of interest. It might further stabilize the results. Application of temporal smoothing should also yield more robust motion estimates, but might result in additional artifacts if the direction of the displacement changes fast. For a sufficiently high temporal resolution it is however a promising direction. Multiscale schemes that exchange information more often between the different scales instead of the used coarse-to-fine-only scheme might avoid being trapped in local minima. And, from a practical point of view, a more efficient implementation of differential operators combined with a tailored stopping criterion for the different scales within the multiscale scheme should be investigated.

5. Joint Motion Estimation and Image Reconstruction

In this chapter, we present a joint image reconstruction and motion estimation scheme. The idea of combining both tasks was first introduced by [138] and recently used in combination with variational schemes by [35, 50, 51]. In comparison to a two-step scheme, i.e., reconstructing images first and then estimating the motion afterwards, we expect higher quality of the reconstructed image sequences and the motion estimates by a joint scheme, as both tasks may endorse each other. In particular, the image reconstruction task and the motion estimation task are solved alternately in our joint approach. This chapter provides the theoretical background for the numerical analysis in Chapter 6. The results were first published in [27]. In [35], the existence of a minimizer in a suitable function space is proven for a 2D setting using the Optical Flow constraint as a motion model. A similar 2D proof for the mass conservation motion model is contained in [50]. In Section 5.1, we extend the setting to the three dimensional case and consider both OF and MC as motion models. Moreover, we extend the analysis to time-dependent forward operators. We propose an unconstrained formulation for an originally constrained problem and prove convergence of the solutions in terms of Γ-convergence. In Section 5.2, we consider the proposed joint approach as a regularization method for a nonlinear ill-posed problem. We state sufficient but restrictive conditions under which the joint approach defines a regularization. Afterwards, we conclude the chapter by describing the numerical scheme used for solving the problem in Section 5.3. We briefly consider the convergence properties of the alternating approach and state algorithmic solutions for both occurring subproblems.

5.1. Existence of a minimizer

We consider a measurable time-dependent image function c on a bounded space-time domain $\Omega \times [0,T] \subset \mathbb{R}^n \times \mathbb{R}^+$, $n \in \{1,2,3\}$, $c \in L^p\big(0,T; BV(\Omega)\big)$, $p > 1$. Functions with bounded variations are a natural choice when describing images, as those functions allow for discontinuities which correspond to edges in the images and they are well-suited to describe homogeneous regions. Having in mind MPI images, both the large homogeneous regions (having in mind the empty background) as well as edges are of importance. The time-dependent image function c has to be reconstructed from measured data $u : [0,T] \to Y$, $u \in L^2(0,T;Y)$ for a reflexive Banach space Y. The data u is obtained by inserting c as an argument into the first component of a linear operator $A : L^{\hat{p}}\left(0,T; L^l(\Omega)\right) \times [0,T] \to L^2(0,T;Y)$ with $\hat{p} = \min(2,p)$ and $l \leq \frac{n}{n-1}$, and corrupting the results with random noise $\delta \in L^2(0, T; Y)$, i.e.,

$$A(c, t) + \delta(t) = u(t), \quad t \in [0, T]. \tag{5.1}$$

Here, we consider $A(\cdot, t) =: A_t(\cdot)$ to be a linear and bounded operator for any $t \in [0, T]$. The considered inverse problem is a dynamic inverse problem for two reasons. First, the image function depends on time. Imagine for example an MPI sequence for tracer agent injected into the bloodstream and propagating through the vascular system. Second, the forward operator is time-dependent, e.g., blurring of dynamic image sequences with time-dependent blur kernels, dynamic CT, or MPI.

In this setting, we want to jointly reconstruct the image function and a velocity field $v : \Omega \times [0, T] \to \mathbb{R}^n$, $v \in L^q\big(0, T; BV(\Omega)^n\big)$, $q > 1$, describing the motion in the data.

In order to handle and stably solve an ill-posed image reconstruction problem, we add a spatial regularizer $R : BV(\Omega) \to \mathbb{R}$, which is assumed to be proper, convex and lower semicontinuous. Moreover, the regularizer on the image function fulfills

$$R(x) \geq |x|_{BV}^p \qquad \text{for any } x \in BV(\Omega). \tag{5.2}$$

As discussed in the previous chapter, the motion estimation task also poses an ill-posed problem. Thus, we add a proper, convex and lower semicontinuous spatial regularizer $S : BV(\Omega)^n \to \mathbb{R}$ working on the motion field. The regularizer is assumed to fulfill the equation

$$S(y) \geq |y|_{BV}^q \qquad \text{for any } y \in BV(\Omega)^n. \tag{5.3}$$

To link the image function c and the motion field v, we introduce a motion model $m(c, v)$ as an additional constraint. The joint image reconstruction and motion estimation problem can then be described by

$$\min_{c,v} \int_0^T D\big(A(c, t), u(t)\big) + \alpha R\big(c(\cdot, t)\big) + \beta S\big(v(\cdot, t)\big) \mathrm{d}t, \tag{5.4}$$

$$\text{s.t. } m(c, v) = 0 \quad \text{in } \mathcal{D}'(\Omega \times [0, T]), \tag{5.5}$$

where $D : Y \times Y \to \mathbb{R}$ denotes the data discrepancy term. We assume D to be proper, convex and lower semicontinuous. Usually, we consider L^1- and L^2-norm differences. As motion models, we consider the optical flow constraint and the mass conservation constraint, i.e.,

$$m_1(c, v) = \frac{\partial}{\partial t} c + \nabla c \cdot v = 0, \tag{5.6}$$

$$m_2(c, v) = \frac{\partial}{\partial t} c + \nabla \cdot (cv) = 0. \tag{5.7}$$

The following theorem shows well-definedness of the problem (5.4), (5.5) for both motion models as solutions exist in suitable function spaces.

Theorem 5.1: Existence of minimizers

Consider the minimization problem (5.4)-(5.5):

$$\min_{c,v} J(c,v) := \int_0^T D\big(A(c,t), u(t)\big) + \alpha R\big(c(\cdot,t)\big) + \beta S\big(v(\cdot,t)\big)\mathrm{d}t,$$
$$\text{s.t. } m(c,v) = 0 \quad \text{in } \mathcal{D}'(\Omega \times [0,T]),$$

in $n \in \{2,3\}$ dimensions.
Let $c \in L^p\big(0,T;BV(\Omega)\big)$, $v \in L^q\big(0,T;BV(\Omega)^n\big)$ and $u \in L^2(0,T;Y)$ for a reflexive Banach space Y over the bounded space-time domain $\Omega \times [0,T]$ and $\Omega \subset \mathbb{R}^n$. Moreover, let $1 < p,q < \infty$ and $\hat{p} = \min(2,p)$. We assume the operator

$$A : L^{\hat{p}}\left(0,T;L^l(\Omega)\right) \times [0,T] \to L^2(0,T;Y)$$

be generated by bounded operators $A_t : L^l(\Omega) \to Y$, $t \in [0,T]$, $l \leq \frac{n}{n-1}$, such that $A_t(c(t)) = A(c,t)$. Further, let $A_t 1_\Omega \neq 0 \quad \forall t \in [0,T]$.
Let R and S fulfill (5.2) and (5.3), i.e.,

$$\begin{aligned} R(x) &\geq |x|^p_{BV} \qquad \text{for any } x \in BV(\Omega), \\ S(y) &\geq |y|^q_{BV} \qquad \text{for any } y \in BV(\Omega)^n. \end{aligned}$$

Moreover, let $D(u_1,u_2) = \frac{1}{2}\|u_1 - u_2\|_Y^2$ be the squared norm distance in Y and m be either m_1 or m_2 as defined in (5.6) and (5.7), respectively. Moreover, assume there exist constants $k_\infty, k_\theta < \infty$ such that $\|v\|_{L^\infty(0,T,L^\infty(\Omega)^n)} \leq k_\infty$ and $\|\nabla \cdot v\|_\theta \leq k_\theta$, where $\theta = L^{p^* s}\left(0,T;L^{l^* k}(\Omega)\right)$ with $1 < s,k < \infty$ and $\frac{1}{p} + \frac{1}{p^*} = 1$.
Then there exists a minimizer of the problem in the set of admissible solutions

$$\Big\{(c,v) \in L^{\hat{p}}\big(0,T;BV(\Omega)\big) \times L^q\big(0,T;BV(\Omega)^n\big) \Big| \; \|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k_\infty,$$
$$\|\nabla \cdot v\|_\theta \leq k_\theta, \;\; m(c,v) = 0 \text{ in } \mathcal{D}'(\Omega \times [0,T])\Big\}. \tag{5.8}$$

Proof. We extend the proof of [35, 50] by several means. First, our image sequence is in up to 3+time dimensions instead of 2+time. The extension of the proof is not completely straightforward as dimension dependent embeddings are used at crucial points and have to be adapted. Second, our regularizers are of a more general form allowing for additional flexibility when applying the theorem. And third, the linear forward operator we consider is time-dependent and has a more general range.
We follow the technique of the direct method in the calculus of variations. The line of argumentation is as follows:

1. The functional $J : L^{\hat{p}}\big(0,T;BV(\Omega)\big) \times L^q\big(0,T;BV(\Omega)^n\big) \to \mathbb{R}$ is bounded from below as all summands under the integral are non-negative. We thus have $\inf J(c,v) > -\infty$. This implies existence of a minimizing sequence of admissible $(c_m,v_m) \in L^{\hat{p}}\big(0,T;BV(\Omega)\big) \times L^q\big(0,T;BV(\Omega)^n\big)$ for $m \in \mathbb{N}$. Admissible means $c_m \in L^p\big(0,T;BV(\Omega)\big)$, $\|v_m\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k_\infty$, $\|\nabla \cdot v_m\|_\theta \leq k_\theta$ and

$m_i(c_m, v_m) = 0$ in $\mathcal{D}'(\Omega \times [0,T])$. We thus have a sequence (c_m, v_m) such that $\lim_{m\to\infty} J(c_m, v_m) = \inf J(c, v)$.

2. We then show that the sublevel sets of J are weak-* compact as follows. As J is continuous, the sublevel sets are closed. It remains to show for this step in the argumentation that the sublevel sets are bounded in norm. This then implies existence of a subsequence $(c_{m_k}, v_{m_k}) \overset{*}{\rightharpoonup} (c^*, v^*)$, which then identifies the candidate (c^*, v^*) for the minimizer.

3. The functional J is lower semi-continuous with respect to the weak-* topology (as a sum of such functionals), which yields

$$\inf J(c, v) \leq J(c^*, v^*) \leq \liminf_{m\to\infty} J(c_{m_k}, v_{m_k}) = \inf J(c, v)\,.$$

4. In order to obtain the weak-* closedness of the admissible set, we prove convergence of the constraint in a distributional sense, i.e., we verify that (c^*, v^*) is admissible. We consider both optical flow and mass conservation constraint and complete the proofs in 3D.

<u>*Step 2, i.e., weak-* compactness of the sublevel sets of J:*</u>
Denote the sublevel set to $\nu \in \mathbb{R}$ by

$$\begin{aligned} S_\nu := \Big\{ (c, v) \in L^{\hat{p}}\big(0, T; BV(\Omega)\big) \times L^q\big(0, T; BV(\Omega)^n\big) : \quad & J(c, v) \leq \nu, \\ & \|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k_\infty \Big\}\,. \end{aligned} \tag{5.9}$$

We will use that $S_\nu \subset L^{\hat{p}}\big(0, T; BV(\Omega)\big) \times L^q\big(0, T; BV(\Omega)^n\big) \cong X^*$ for X^* being the dual space of a Banach space X. Therefore S_ν is weak-* compact if and only if S_ν is closed in weak-* topology and bounded in norm (cf. Theorem 2.4). The closedness of the sublevel set holds as J is continuous. It remains to show the boundedness in norm, i.e., it remains to show that

$$\begin{aligned} \|c\|_{L^{\hat{p}}(0,T;BV(\Omega))} = \left(\int_0^T \Big(\|c\|_{L^1(\Omega)} + |c|_{BV(\Omega)} \Big)^{\hat{p}} \mathrm{d}t \right)^{1/\hat{p}} &\leq k, \\ \|v\|_{L^q(0,T;BV(\Omega)^n)} &\leq k \quad \text{for a} \quad k < \infty. \end{aligned}$$

Let $(c, v) \in S_\nu$ (i.e, it holds additionally $\|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k_\infty$).
We start with the bound on the norm of c: From $(c, v) \in S_\nu$ it follows that

$$\begin{aligned} \int_0^T \frac{1}{2} \|(A_t c)(t) - u(t)\|_Y^2 \,\mathrm{d}t = \frac{1}{2} \|A_t c - u\|_Y^2 \leq \nu \\ \Rightarrow (A_t c - u) \in Y \quad \text{a. e. in} \quad [0, T]\,. \end{aligned}$$

Define $k_A(t) := \|(A_t c)(t) - u(t)\|_Y$, then

$$\int_0^T k_A^{\hat{p}} \mathrm{d}t = \int_0^T \|A_t c - u\|_Y^{\hat{p}} \,\mathrm{d}t$$

is bounded for $1 \leq \hat{p} \leq 2$ due to the continuous embedding $L^2(0,T;Y) \hookrightarrow L^{\hat{p}}(0,T;Y)$ (see Lemma 2.3).
To bound $\|c\|_{L^{\hat{p}}(0,T;L^1(\Omega))}$, we start by an upper bound for arbitrary but fixed $t \in [0,T]$. Defining the average $\bar{c} = \frac{1}{|\Omega|}\int_\Omega c(x,t)\,\mathrm{d}x$ and $c_0 = c(\cdot,t) - \bar{c}$, we note that

$$\int_\Omega c_0 \mathrm{d}x = 0.$$

Moreover,

$$|c_0|_{BV(\Omega)} = |c(\cdot,t)|_{BV(\Omega)} \leq k \int_0^T |c(\cdot,\tau)|^p_{BV(\Omega)}\,\mathrm{d}\tau \leq \frac{k\nu}{\alpha},$$

for a constant $k \in \mathbb{R}$. With Poincaré-Wirtinger, it follows that there exists a constant $k_1 > 0$ such that

$$\|c_0\|_{L^l(\Omega)} \leq k_1 |c|_{BV(\Omega)} \leq k_1 \frac{k\nu}{\alpha} =: \tilde{k}_1 \tag{5.10}$$

for $l \leq \frac{n}{n-1}$. We derive the following estimate:

$$\begin{aligned}
&\|A_t\bar{c}\|_Y^2 - 2\|A_t\bar{c}\|_Y \Big(\|A_t\|\,\|c_0\|_{L^l(\Omega)} + \|u\|_Y\Big)\\
&\quad\leq \|A_t\bar{c}\|_Y^2 - 2\|A_tc_0 - u\|_Y \|A_t\bar{c}\|_Y\\
&\quad\leq \|A_tc_0 - u\|_Y^2 + \|A_t\bar{c}\|_Y^2 - 2\|A_tc_0 - u\|_{L^2(Y)} \|A_t\bar{c}\|_Y\\
&\quad= \big(\|A_tc_0 - u\|_Y - \|A_t\bar{c}\|_Y\big)^2\\
&\quad\leq \|A_tc_0 - u + A_t\bar{c}\|_Y^2\\
&\quad= \|A_t(c_0 + \bar{c}) - u\|_Y^2 = k_A^2(t)
\end{aligned}$$

Using the estimate of (5.10) and $u \in L^2(0,T;Y)$ yields

$$0 \leq \|A_t\|\,\|c_0\|_{L^l(\Omega)} + \|u\|_Y \leq \|A_t\|\,\tilde{k}_1 + \|u\|_Y := k_2(t).$$

Note that, as $\|A_t\|$ is bounded for all $t \in [0,T]$, we also have a bound on the time integral $\int_0^T k_2(t)\mathrm{d}t$. We can now combine the estimates above to obtain a bound on $\|A_t\bar{c}\|_Y$:

$$\begin{aligned}
&\|A_t\bar{c}\|_Y^2 - 2\|A_t\bar{c}\|_Y \Big(\|A_t\|\,\|c_0\|_{L^l(\Omega)} + \|u\|_Y\Big)\\
&\quad + \Big(\|A_t\|\,\|c_0\|_{L^l(\Omega)} + \|u\|_Y\Big)^2 \leq k_A^2(t) + k_2^2(t)\\
&\quad\Leftrightarrow \Big(\|A_t\bar{c}\|_Y - \|A_t\|\,\|c_0\|_{L^l(\Omega)} + \|u\|_Y\Big)^2 \leq k_A^2(t) + k_2^2(t)\\
&\quad\Rightarrow \|A_t\bar{c}\|_Y \leq \big(k_A^2(t) + k_2^2(t)\big)^{1/2} + \|A_t\|\,\|c_0\|_{L^l(\Omega)} + \|u\|_Y\\
&\quad\leq \big(k_A^2(t) + k_2^2(t)\big)^{1/2} + k_2(t) := k_3(t)
\end{aligned}$$

Finally, we deduce a bound on $\|c(\cdot,t)\|_{L^1(\Omega)}$ by

$$\begin{aligned}
0 &\leq \|c(\cdot,t)\|_{L^1(\Omega)} \leq k_4 \|c(\cdot,t)\|_{L^l(\Omega)} = k_4\|c_0 + \bar{c}\|_{L^l(\Omega)}\\
&\leq k_4 \Big(\|c_0\|_{L^l(\Omega)} + |\bar{c}|\,\|\chi_\Omega\|_{L^l(\Omega)}\Big) \leq k_4 \left(\tilde{k}_1 + \frac{k_3(t)\,\|\chi_\Omega\|_{L^l(\Omega)}}{\|A_t\chi_\Omega\|_Y}\right),
\end{aligned}$$

where we used that by assuming $\|A_t\chi_\Omega\|_Y \neq 0$ it holds that

$$|\bar{c}| \, \|A_t\chi_\Omega\|_Y = \|A_t\bar{c}\|_Y \leq k_3(t)$$

and thus

$$|\bar{c}| \leq \frac{k_3(t)}{\|A_t\chi_\Omega\|_Y}.$$

We now have

$$\int_0^T \|c(\cdot,t)\|_{L^1(\Omega)}^{\hat{p}} \,\mathrm{d}t \leq \int_0^T \left(k_4\left(\tilde{k}_1 + \frac{k_3(t)\,\|\chi_\Omega\|_{L^l(\Omega)}}{\|A_t\chi_\Omega\|_Y}\right)\right)^{\hat{p}} \mathrm{d}t \leq: k_M < \infty,$$

for $1 \leq \hat{p} \leq 2$, where k_M exists as the expression is bounded for all t and thus the supremum is also bounded. Putting it all together, we arrive with $j := p/\hat{p}$ at

$$\begin{aligned}
\|c\|_{L^{\hat{p}}(0,T;BV(\Omega))}^{\hat{p}} &= \int_0^T \|c\|_{BV(\Omega)}^{\hat{p}} \,\mathrm{d}t \\
&\leq 2^{\hat{p}-1}\left(\int_0^T |c|_{BV(\Omega)}^{\hat{p}} \,\mathrm{d}t + \int_0^T \|c\|_{L^1(\Omega)}^{\hat{p}} \,\mathrm{d}t\right) \\
&\leq 2^{\hat{p}-1}\int_0^T |c|_{BV(\Omega)}^{\hat{p}} \,\mathrm{d}t + 2^{\hat{p}-1}k_M \\
&\leq 2^{\hat{p}-1}\left(\int_0^T \left|\chi_{[0,T]}\right|^{j^*} \mathrm{d}t\right)^{\frac{1}{j^*}} \left(\int_0^T |c|_{BV(\Omega)}^{\hat{p}j} \,\mathrm{d}t\right)^{\frac{1}{j}} + 2^{\hat{p}-1}k_M \\
&= 2^{\hat{p}-1}k_M + 2^{\hat{p}-1}\left\|\chi_{[0,T]}\right\|_{L^{j^*}(0,T)} \left(\int_0^T |c|_{BV(\Omega)}^{p} \,\mathrm{d}t\right)^{\frac{1}{j}} \\
&\leq 2^{\hat{p}-1}\left(k_M + \left\|\chi_{[0,T]}\right\|_{L^{j^*}(0,T)}\left(\frac{\nu}{\alpha}\right)^{\frac{1}{j}}\right) < \infty,
\end{aligned}$$

where $1/j + 1/j^* = 1$.
Consider now the norm of v: By assumption, we know that there exists k_∞ such that $\|v\|_{L^\infty(\Omega)^n} \leq k_\infty$ almost everywhere on $[0,T]$. We proceed by the following estimate:

$$\begin{aligned}
\|v\|_{L^q(0,T;BV(\Omega)^n)}^q &= \int_0^T \left(\|v\|_{L^1(\Omega)^n} + |v|_{BV(\Omega)^n}\right)^q \mathrm{d}t \\
&\leq 2^{q-1}\left(\int_0^T \|v\|_{L^1(\Omega)^n}^q \,\mathrm{d}t + \int_0^T |v|_{BV(\Omega)^n}^q \,\mathrm{d}t\right) \\
&\leq 2^{q-1}\left(\int_0^T \left(k_\infty\,|\Omega|\right)^q \mathrm{d}t + \frac{\nu}{\beta}\right) \\
&= 2^{q-1}\left(T\left(k_\infty\,|\Omega|\right)^q + \frac{\nu}{\beta}\right) < \infty.
\end{aligned}$$

For the second inequality, we used that $v \in S_\nu$ and thus $\int_0^T \beta\,|v|_{BV(\Omega)^n}^q \,\mathrm{d}t \leq \nu$ by assumption.
We conclude that the sublevel sets S_ν are bounded in norm for fixed $\nu \in \mathbb{R}$, as both $\|v\|_{L^q(0,T;BV(\Omega)^n)}$ and $\|c\|_{L^{\hat{p}}(0,T;BV(\Omega))}$ are bounded and the bounds depend on ν only.

Moreover, we know that $BV(\Omega)$ is isometrically isomorphic to the dual space of some Banach space Z (cf. Lemma 2.4) and thus

$$L^{\hat{p}}\big(0,T;BV(\Omega)\big) \cong \big(L^{p^*}(0,T;Z)\big)^* \quad \text{with} \quad \frac{1}{\hat{p}}+\frac{1}{p^*}=1,$$
$$L^{q}\big(0,T;BV(\Omega)^n\big) \cong \big(L^{q^*}(0,T;Z^n)\big)^* \quad \text{with} \quad \frac{1}{q}+\frac{1}{q^*}=1.$$

Consequently, both spaces are dual spaces and thus

$$S_\nu \subset L^{\hat{p}}\big(0,T;BV(\Omega)\big) \times L^q\big(0,T;BV(\Omega)^n\big) \cong X^*$$

for X^* being the dual of a Banach space X (here $X = L^{p^*}(0,T;Z) \times L^{q^*}(0,T;Z^n)$). Moreover, S_ν is closed in weak-* topology as J is continuous and the sublevel sets are bounded in norm. It follows that S_ν is weak-* compact by Banach-Alaoglu (cf. Theorem 2.4). This implies existence of a subsequence $(c_{m_k}, v_{m_k}) \xrightarrow{*} (c^*, v^*)$, which then identifies the candidate (c^*, v^*) for the minimizer.

Step 3, i.e., lower semicontinuity of J with respect to the weak- topology:*
As a sum of weak-* lower semicontinuous functionals, J itself is weak-* lower semicontinuous, cf. [50].

Step 4a, i.e., convergence of the constraint for the optical flow constraint:
Let $(c_m, v_m) \in L^{\hat{p}}\big(0,T;BV(\Omega)\big) \times L^q\left(0,T;BV(\Omega)^n\right), m \in \mathbb{N}$, be an admissible sequence (i.e., $c_m \in L^p\big(0,T,BV(\Omega)\big)$, $\|v_m\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \le k_\infty$, $\|\nabla \cdot v_m\|_\theta \le k_\theta$ and $m_1(c,v) = 0$ in $\mathcal{D}'(\Omega \times [0,T])$), which also fulfills $(c_m, v_m) \in S_\nu$ for some $\nu \in \mathbb{R}$. Then c_m and v_m are bounded and it exist c and v such that by passing over to a subsequence (again denoted by c_m and v_m) we have

$$c_m \xrightarrow{*} c, \quad v_m \xrightarrow{*} v.$$

We want to show that

$$\frac{\partial}{\partial t} c_m + \nabla c_m \cdot v_m \longrightarrow \frac{\partial}{\partial t} c + \nabla c \cdot v \quad \text{in } \mathcal{D}'(\Omega \times [0,T]),$$

i.e., we have convergence of the constraint in a distributional sense. Therefore we need strong convergence of at least one factor of $\nabla c_m \cdot v_m$ in a certain sense. In order to use the Lemma of Aubin-Lions, we need boundedness of the time derivative of the sequence c_m as well as some specific embeddings. We start with the bound for $\frac{\partial}{\partial t} c$:

$$\left| \int_0^T \int_\Omega \frac{\partial}{\partial t} c\varphi \; \mathrm{d}x\mathrm{d}t \right| = \left| \int_0^T \int_\Omega c \nabla \cdot (v\varphi) \, \mathrm{d}x\mathrm{d}t \right| \quad \forall \varphi \in \mathcal{C}_C^\infty\big(\Omega \times (0,T)\big)$$
$$\le \int_0^T \|c\nabla \cdot (v\varphi)\|_{L^1(\Omega)} \, \mathrm{d}t,$$
$$\le \int_0^T \|c\|_{L^l(\Omega)} \|\nabla \cdot (v\varphi)\|_{L^{l^*}(\Omega)} \, \mathrm{d}t,$$

$$
\begin{aligned}
&\leq \int_0^T \|c\|_{L^l(\Omega)} \left(\|\varphi \nabla \cdot v\|_{L^{l^*}(\Omega)} + \|v \cdot \nabla \varphi\|_{L^{l^*}(\Omega)} \right) \mathrm{d}t, \\
&= \underbrace{\int_0^T \|c\|_{L^l(\Omega)} \|\varphi \nabla \cdot v\|_{L^{l^*}(\Omega)} \,\mathrm{d}t}_{(I)} + \underbrace{\int_0^T \|c\|_{L^l(\Omega)} \|v \cdot \nabla \varphi\|_{L^{l^*}(\Omega)} \,\mathrm{d}t}_{(II)},
\end{aligned}
$$

where the second inequality results from applying Hölders inequality with $1/l + 1/l^* = 1$ for $l \leq n/(n-1)$ and for the third inequality, we applied the Minkowksi inequality. The first term (I) can be estimated by

$$
\begin{aligned}
&\int_0^T \|c\|_{L^l(\Omega)} \|\varphi \nabla \cdot v\|_{L^{l^*}(\Omega)} \,\mathrm{d}t = \left\| \|c\|_{L^l(\Omega)} \|\varphi \nabla \cdot v\|_{L^{l^*}(\Omega)} \right\|_{L^1(0,T)} \\
&\leq \|c\|_{L^p(0,T;L^l(\Omega))} \|\varphi \nabla \cdot v\|_{L^{p^*}(0,T;L^{l^*}(\Omega))}, \\
&= \|c\|_{L^p(0,T;L^l(\Omega))} \left\| \|\varphi \nabla \cdot v\|_{L^{l^*}(\Omega)} \right\|_{L^{p^*}(0,T)} \\
&\leq k_c \left\| \|\varphi\|_{L^{l^*k^*}(\Omega)} \|\nabla \cdot v\|_{L^{l^*k}(\Omega)} \right\|_{L^{p^*}(0,T)}, \\
&\leq k_c \|\varphi\|_{L^{p^*s^*}(0,T;L^{l^*k^*}(\Omega))} \|\nabla \cdot v\|_{L^{p^*s}(0,T;L^{l^*k}(\Omega))}, \\
&\leq k_c \|\varphi\|_{L^{p^*s^*}(0,T;L^{l^*k^*}(\Omega))} k_\theta,
\end{aligned}
$$

using Hölder with $1/p + 1/p^* = 1$ for $1 < p \leq 2$ for the first inequality, Hölder with $1/(l^*k) + 1/(l^*k^*) = 1/l^*$ for $1 \leq k \leq \infty$ for the second inequality, and Hölder with $1/(p^*s) + 1/(p^*s^*) = 1/p^*$ for $1 \leq s < \infty$ for the third inequality. Additionally, we used that c is bounded in $L^p\big(0,T;BV(\Omega)\big)$ by assumption and as $BV(\Omega) \hookrightarrow L^l(\Omega)$ for $l \leq n/(n-1)$ (cf. Theorem 2.2) there exists a constant k_c such that $\|c\|_{L^p(0,T;L^l(\Omega))} \leq k_c$. For the second term, namely (II), we derive a bound by

$$
\begin{aligned}
&\int_0^T \|c\|_{L^l(\Omega)} \|v \cdot \nabla \varphi\|_{L^{l^*}(\Omega)} \,\mathrm{d}t \qquad (5.11)\\
&\leq \int_0^T \|c\|_{L^l(\Omega)} \||v| \cdot |\nabla \varphi|\|_{L^{l^*}(\Omega)} \,\mathrm{d}t, \\
&\leq \int_0^T \|c\|_{L^l(\Omega)} \|v\|_{L^\infty(\Omega)^n} \|\nabla \varphi\|_{L^{l^*}(\Omega)^n} \,\mathrm{d}t \\
&\leq \int_0^T \|c\|_{L^l(\Omega)} k_\infty \|\varphi\|_{W^{1,l^*}(\Omega)} \,\mathrm{d}t \\
&\leq k_\infty \|c\|_{L^p(0,T;L^l(\Omega))} \|\varphi\|_{L^{p^*}(0,T;W^{1,l^*}(\Omega))}, \\
&\leq k_\infty k_c \|\varphi\|_{L^{p^*}(0,T;W^{1,l^*}(\Omega))}, \qquad (5.12)
\end{aligned}
$$

using Hölder with $1/p + 1/p^* = 1$ for $1 < p \leq 2$ for the fourth inequality and Cauchy-Schwarz for the first inequality. Combining the results for (I) and (II) yields

$$
\begin{aligned}
\left| \int_0^T \int_\Omega \frac{\partial}{\partial t} c\varphi \,\mathrm{d}x\mathrm{d}t \right| &\leq k_c k_\theta \|\varphi\|_{L^{p^*s^*}(0,T;L^{l^*k^*}(\Omega))} + k_\infty k_c \|\varphi\|_{L^{p^*}(0,T;W^{1,l^*}(\Omega))} \\
&\leq k_c \left(k_\theta \|\varphi\|_{L^{p^*s^*}(0,T;L^{l^*k^*}(\Omega))} + k_\infty k_6 \|\varphi\|_{L^{p^*s^*}(0,T;W^{1,l^*}(\Omega))} \right).
\end{aligned}
$$

The last estimate was obtained by using $s^* > 1$ and thus $L^{p^*s^*}(\Omega) \hookrightarrow L^{p^*}(\Omega)$ (see Lemma 2.3), which defines the constant k_6. We now use the dimension dependent Sobolev embedding (cf. Theorem 2.1)

$$W^{1,l^*}(\Omega) \hookrightarrow L^{l^*k^*}(\Omega),$$

which exists for all $1 \leq k^* < \infty$, as $l^* \geq n$ and then $l^* \leq l^*k^* < \infty$.
This yields

$$\begin{aligned}\left|\int_0^T \int_\Omega \frac{\partial}{\partial t} c\varphi \,\mathrm{d}x\mathrm{d}t\right| &\leq k_c \left(k_\theta \|\varphi\|_{L^{p^*s^*}(0,T;W^{1,l^*}(\Omega))} + k_\infty k_6 \|\varphi\|_{L^{p^*s^*}(0,T;W^{1,l^*}(\Omega))}\right) \\ &= \|\varphi\|_{L^{p^*s^*}(0,T;W^{1,l^*}(\Omega))} \left(k_c k_\theta + k_c k_\infty k_6\right).\end{aligned}$$

We conclude that the temporal derivative $\frac{\partial}{\partial t}c$ acts as a bounded linear functional on $L^{p^*s^*}\left(0,T;W_0^{1,l^*}(\Omega)\right)$ and thus

$$\frac{\partial}{\partial t}c \in \left(L^{p^*s^*}\left(0,T;W_0^{1,l^*}(\Omega)\right)\right)^* = L^{\frac{ps}{p+s-1}}\left(0,T;W^{-1,l}(\Omega)\right).$$

We now use the Lemma of Aubin-Lions [11, 103], cf. Theorem 2.3, identifying the Banach spaces as $X = BV(\Omega)$, $Y = L^r(\Omega)$ and $Z = W^{-1,l}(\Omega)$ such that $X \Subset Y$, $Y \hookrightarrow Z$, c_i a sequence of bounded functions in $L^p(0,T;X)$ and $\frac{\partial}{\partial t}cc_i$ bounded in $L^q(0,T;Z)$ with either $q = 1$ and $1 \leq p < \infty$ or $q > 1$ and $1 \leq p \leq \infty$. Then c_i is relatively compact in $L^p(0,T;Y)$. The embedding $BV(\Omega) \Subset L^r(\Omega)$ holds for $r < \frac{n}{n-1}$ [5], i.e.

$$\begin{cases} r < 2, & n = 2 \\ r < 1.5, & n = 3. \end{cases}.$$

It remains to identify for which r the continuous embedding $L^r(\Omega) \hookrightarrow W^{-1,l}(\Omega)$ exists. The embedding holds if $W^{1,l^*}(\Omega) \hookrightarrow L^{r^*}(\Omega)$, which by the Sobolev embedding theorem for the different dimensions and $l^* \geq n$ translates to

$$W^{1,l^*}(\Omega) \hookrightarrow L^{r^*}(\Omega) \text{ for } l^* \leq r^* < \infty. \tag{5.13}$$

In terms of r, this results in $1 < r \leq l$.
Applying Aubin-Lions yields $\left\{c \in L^p\big(0,T;BV(\Omega)\big) \middle| \exists k > 0 \text{ s.t. } \|c\|_{L^p(0,T;BV(\Omega))} \leq k,\right.$ $\left.\frac{\partial}{\partial t}c + v\cdot\nabla c = 0 \text{ in } \mathcal{D}'(\Omega\times[0,T])\right\}$ is relatively compact in $L^p\big(0,T;L^r(\Omega)\big)$ for r satisfying

$$\begin{cases} 1 \leq r < 2, & n = 2 \\ 1 \leq r < \frac{3}{2}, & n = 3 \end{cases}. \tag{5.14}$$

The sequence c_m thus converges even strongly to c in $L^p\big(0,T;L^r(\Omega)\big)$.
We are now settled to prove convergence of the constraint.
In the following, let $\varphi \in \mathcal{C}_0^\infty\big(\Omega\times[0,T]\big)$. We start with the time derivative:

$$\int_0^T \int_\Omega \left(\left(\frac{\partial}{\partial t}c\right)_m - \frac{\partial}{\partial t}c\right)\varphi \,\mathrm{d}x\mathrm{d}t = -\int_0^T \int_\Omega (c_m - c)\frac{\partial}{\partial t}\varphi \,\mathrm{d}x\mathrm{d}t,$$

by integration by parts.
We know that $c_m \overset{*}{\rightharpoonup} c$ in $L^p\big(0,T;BV(\Omega)\big) \cong \big(L^{p^*}(0,T;Z)\big)^*$ for $BV(\Omega)$ being isometrically isomorphic to the dual space of Z. As $\varphi \in \mathcal{C}_0^\infty\big(\Omega \times [0,T]\big)$, it follows that $\frac{\partial}{\partial t}\varphi \in L^{p^*}(0,T;Z)$ and this yields

$$-\int_0^T \int_\Omega (c_m - c)\,\frac{\partial}{\partial t}\varphi \,\mathrm{d}x\mathrm{d}t \longrightarrow 0 \quad \text{for } m \longrightarrow \infty.$$

For the product term, it holds that

$$\begin{aligned}
&-\int_0^T \int_\Omega (\nabla c_m \cdot v_m - \nabla c \cdot v)\,\varphi \,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_\Omega c_m \nabla\cdot(v_m\varphi) - c\nabla\cdot(v\varphi)\,\mathrm{d}x\,\mathrm{d}t\\
&= \underbrace{\int_0^T \int_\Omega (c_m - c)\,\nabla\cdot(v_m\varphi)\,\mathrm{d}x\,\mathrm{d}t}_{(I)} + \underbrace{\int_0^T \int_\Omega c\nabla\cdot(\varphi v_m - \varphi v)\,\mathrm{d}x\,\mathrm{d}t}_{(II)}.
\end{aligned}$$

For (I) we obtain

$$\begin{aligned}
&\int_0^T \int_\Omega (c_m - c)\,\nabla\cdot(v_m\varphi)\,\mathrm{d}x\mathrm{d}t\\
&\le \int_0^T \|c_m - c\|_{L^r(\Omega)}\,\|\nabla\cdot(v_m\varphi)\|_{L^{r^*}(\Omega)}\,\mathrm{d}t\\
&\le \|c_m - c\|_{L^p(0,T;L^r(\Omega))}\,\|\nabla\cdot(v_m\varphi)\|_{L^{p^*}(0,T;L^{r^*}(\Omega))}\\
&\le \|c_m - c\|_{L^p(0,T;L^r(\Omega))}\Big(\|\nabla\varphi\cdot v_m\|_{L^{p^*}(0,T;L^{r^*}(\Omega))} + \|\varphi\nabla\cdot(v_m)\|_{L^{p^*}(0,T;L^{r^*}(\Omega))}\Big) \qquad (5.15)\\
&\le \|c_m - c\|_{L^p(0,T;L^r(\Omega))}\Big(\underbrace{\|\nabla\varphi\cdot v_m\|_{L^{p^*}(0,T;L^{r^*}(\Omega))}}_{(a)}\\
&\quad + \underbrace{\|\varphi\|_{L^{p^*s^*}(0,T;L^{r^*}(\Omega))}}_{(b)}\underbrace{\|\nabla\cdot(v_m)\|_{L^{p^*s}(0,T;L^{r^*}(\Omega))}}_{(c)}\Big), \qquad (5.16)
\end{aligned}$$

using Hölder with $1/r + 1/r^* = 1$ for $1 < r < \infty$, $1/p + 1/p^* = 1$ for $1 < p \le 2$ and $1/(p^*s) + 1/(p^*s^*) = 1/p^*$ for $1 < s < \infty$.
We first note that (b) is bounded as φ is a test function. Consider now (c):
$\|\nabla\cdot(v_m)\|_{L^{p^*s}(0,T;L^{l^*k}(\Omega))}$ is bounded by k_θ by assumption. We thus need an embedding

$$L^{p^*s}\big(0,T;L^{l^*k}(\Omega)\big) \hookrightarrow L^{p^*s}\big(0,T;L^{r^*}(\Omega)\big)$$

and therefore $r^* \le l^*k$, which is equivalent to $r \ge \frac{l^*k}{l^*k-1}$. We need to adjust the bounds on k as follows.

- Case $n = 2$: We have $1 \le r < 2$ and $l^* \ge 2$. We note that $1 \le \frac{l^*k}{l^*k-1}$ is always fulfilled, but in order to guarantee $\frac{l^*k}{l^*k-1} < 2$, we have to restrict k such that $\big[\frac{l^*k}{l^*k-1}, 2\big) \ne \emptyset$. This is ensured by $l^*k > 2$, which is equivalent to $k > 1$, as $l^* \ge 2$.
- Case $n = 3$: We have $1 \le r < \frac{3}{2}$ and $l^* \ge 3$. We note that $1 \le \frac{l^*k}{l^*k-1}$ is always fulfilled, but in order to guarantee $\frac{l^*k}{l^*k-1} < \frac{3}{2}$, we have to restrict k such that $\big[\frac{l^*k}{l^*k-1}, \frac{3}{2}\big) \ne \emptyset$. This is ensured by $l^*k > 3$, which is equivalent to $k > 1$, as $l^* \ge 3$.

It remains to show that (a) is bounded. This holds as φ is a test function and thus $\nabla\varphi$ is bounded and $\|v\|_{L^\infty(\Omega)^n} \leq k_\infty$. We showed that $(I) \longrightarrow 0$ as $m \longrightarrow \infty$. Now consider (II), i.e.

$$\int_0^T \int_\Omega c\nabla \cdot (\varphi v_m - \varphi v) \,\mathrm{d}x \,\mathrm{d}t \tag{5.17}$$

$$= \underbrace{\int_0^T \int_\Omega c\varphi\nabla \cdot (v_m - v) \,\mathrm{d}x \,\mathrm{d}t}_{(a)} + \underbrace{\int_0^T \int_\Omega c\,(v_m - v) \cdot \nabla\varphi \,\mathrm{d}x\mathrm{d}t}_{(b)}. \tag{5.18}$$

Consider (a) first. By assumption, $\nabla \cdot (v_m - v) \in L^{p^*s}\big(0,T; L^{l^*k}(\Omega)\big)$. We need to show $c\varphi \in \big(L^{p^*s}\big(0,T;L^{l^*k}(\Omega)\big)\big)^*$. Starting from $c \in L^p\big(0,T;BV(\Omega)\big)$, the first embedding needed is $BV(\Omega) \hookrightarrow L^{\frac{l^*k}{l^*k-1}}(\Omega) = L^{(l^*k)^*}(\Omega)$.

- Case $n = 2$: $BV(\Omega) \hookrightarrow L^{\frac{l^*k}{l^*k-1}}(\Omega)$ exists continuously for $1 \leq \frac{l^*k}{l^*k-1} \leq 2$. This is fulfilled for $l^*k \geq 2$, which is ensured by $l^* \geq 2$ and $k > 1$.

- Case $n = 3$: $BV(\Omega) \hookrightarrow L^{\frac{l^*k}{l^*k-1}}(\Omega)$ exists for $1 \leq \frac{l^*k}{l^*k-1} \leq 1.5$. Again, this holds for $l^*k \geq 3$, which is ensured by $l^* \geq 3$ and $k > 1$.

Note that $(p^*s)^* = \frac{ps}{ps-p+1}$ and $\frac{ps}{ps-p+1} < p$ if $1 < p$, which again holds by assumption. In summary, we have

$$L^p\big(0,T;BV(\Omega)\big) \hookrightarrow L^p\left(0,T;L^{(l^*k)^*}(\Omega)\right) \hookrightarrow L^{(p^*s)^*}\left(0,T;L^{(l^*k)^*}(\Omega)\right).$$

As φ and its derivatives are bounded, $c\varphi \in L^{(p^*s)^*}\big(0,T;L^{(l^*k)^*}(\Omega)\big)$. Moreover, from $v_m \overset{*}{\rightharpoonup} v$ it follows that $\nabla \cdot v_m \overset{*}{\rightharpoonup} \nabla \cdot v$ and thus altogether

$$\int_0^T \int_\Omega c\varphi\nabla \cdot (v_m - v) \,\mathrm{d}x\mathrm{d}t \longrightarrow 0 \;\text{ for }\; m \longrightarrow \infty.$$

Considering (b), we denote first that

$$BV\,(\Omega) \hookrightarrow L^1\,(\Omega)$$

and thus

$$L^p\big(0,T;BV(\Omega)\big) \hookrightarrow L^p\left(0,T;L^1(\Omega)\right) \hookrightarrow L^1\left(0,T;L^1(\Omega)\right).$$

Therefore, we have $c \in L^1\big(0,T;L^1(\Omega)\big)$ and by the boundedness of φ and its derivatives, we obtain $c\nabla\varphi \in L^1\big(0,T;L^1(\Omega)^n\big)$. Moreover, by the bound $\|v\|_{L^\infty(\Omega)^n} \leq k_\infty$ we see that $(v_m - v) \in L^\infty\big(0,T;L^\infty(\Omega)^n\big)$. Now by

$$L^\infty\big(0,T;L^\infty(\Omega)^n\big) \cong \Big(L^1\big(0,T;L^1(\Omega)^n\big)\Big)^*$$

and the weak-* convergence of v_m to v, we reach

$$\int_0^T \int_\Omega c\nabla\varphi \cdot (v_m - v) \,\mathrm{d}x\mathrm{d}t \to 0 \;\text{ for }\; m \to \infty.$$

Combining all the results yields the convergence of the constraint

$$\int_0^T \int_\Omega \left(\left(\left(\frac{\partial}{\partial t} c \right)_m - \nabla c_m \cdot v_m \right) - \left(\frac{\partial}{\partial t} c - \nabla c \cdot v \right) \right) \varphi \,\mathrm{d}x\mathrm{d}t \to 0 \text{ for } m \to \infty.$$

<u>*Step 4b, i.e., convergence of the constraint for the mass conservation constraint*:</u>
Let $(c_m, v_m) \in L^{\hat{p}}\big(0,T; BV(\Omega)\big) \times L^q\big(0,T; BV(\Omega)^n\big), m \in \mathbb{N}$, be an admissible sequence (i.e., $c_m \in L^p\big(0,T,BV(\Omega)\big)$, $\|v_m\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k_\infty$, $\|\nabla \cdot v_m\|_\theta \leq k_\theta$ and $m_2(c,v) = 0$ in $\mathcal{D}'(\Omega \times [0,T])$), which also fulfills $(c_m, v_m) \in S_\nu$ for some $\nu \in \mathbb{R}$. Then c_m and v_m are bounded and it exist c and v such that by passing over to a subsequence (again denoted by c_m and v_m) we have

$$c_m \overset{*}{\rightharpoonup} c, \quad v_m \overset{*}{\rightharpoonup} v.$$

We want to show that

$$\left(\frac{\partial}{\partial t} c \right)_m + \nabla \cdot (c_m v_m) \longrightarrow \frac{\partial}{\partial t} c + \nabla \cdot (cv) \quad \text{in } \mathcal{D}'(\Omega \times [0,T]),$$

i.e., we have convergence of the constraint in a distributional sense. We start again with a bound on $\frac{\partial}{\partial t}c$:

$$\begin{aligned}
\left| \int_0^T \int_\Omega \frac{\partial}{\partial t} c \varphi \,\mathrm{d}x\mathrm{d}t \right| &= \left| \int_0^T \int_\Omega -\nabla \cdot (cv)\, \varphi \,\mathrm{d}x\mathrm{d}t \right| = \left| \int_0^T \int_\Omega (cv) \cdot \nabla \varphi \,\mathrm{d}x\mathrm{d}t \right| \\
&\leq \int_0^T \int_\Omega |(cv) \cdot \nabla \varphi| \,\mathrm{d}x\mathrm{d}t \\
&\leq \int_0^T \|c\|_{L^l(\Omega)} \|v \cdot \nabla \varphi\|_{L^{l^*}(\Omega)} \,\mathrm{d}t, \\
&\leq k_\infty k_c \|\varphi\|_{L^{p^*}(0,T;W^{1,l^*}(\Omega))}, \quad \text{(c.f. (5.11) – (5.12))},
\end{aligned} \tag{5.19}$$

using Hölder with $1/l + 1/l^* = 1$ for $l \leq n/(n-1)$. We thus know that $\frac{\partial}{\partial t}c$ acts as a bounded linear functional on $L^{p^*}\left(0,T; W^{1,l^*}(\Omega)\right)$ and thus $\frac{\partial}{\partial t}c \in L^p\left(0,T; W^{-1,l}(\Omega)\right)$. The Lemma of Aubin-Lions can be applied similarly to the optical flow case and yields strong convergence of $c_m \longrightarrow c$ in $L^p(0,T; L^r(\Omega))$ with the same constraints on r. Also, the arguments for the convergence of the time derivative

$$-\int_0^T \int_\Omega (c_m - c) \frac{\partial}{\partial t} \varphi \,\mathrm{d}x\mathrm{d}t \longrightarrow 0 \quad \text{for } m \longrightarrow \infty.$$

are the same as in the optical flow case. We now come to the product term. It remains to show that

$$\nabla \cdot (c_m v_m) \rightharpoonup \nabla \cdot (cv).$$

$$\begin{aligned}
-\int_0^T \int_\Omega (\nabla \cdot (c_m v_m) - \nabla \cdot (cv))\, \varphi \,\mathrm{d}x\mathrm{d}t &= \int_0^T \int_\Omega (c_m v_m - cv \pm c v_m) \cdot \nabla \varphi \,\mathrm{d}x\mathrm{d}t \\
&= \int_0^T \int_\Omega ((c_m - c)\, v_m + c\,(v_m - v)) \cdot \nabla \varphi \,\mathrm{d}x\,\mathrm{d}t
\end{aligned}$$

$$
\begin{aligned}
&\leq \int_0^T \|c_m - c\|_{L^r(\Omega)} \|v_m \cdot \nabla\varphi\|_{L^{r^*}(\Omega)} + \int_\Omega c\,(v_m - v)\cdot \nabla\varphi \,\mathrm{d}x\,\mathrm{d}t,\\
&\leq \underbrace{\|c_m - c\|_{L^p(0,T;L^r(\Omega))} \|v_m \cdot \nabla\varphi\|_{L^{p^*}(0,T;L^{r^*}(\Omega))}}_{\longrightarrow 0 \quad \text{for} \quad m \longrightarrow \infty \quad (\text{cf.} \quad (5.16)(a))}\\
&+ \underbrace{\int_0^T \int_\Omega c\,(v_m - v)\cdot\nabla\varphi\,\mathrm{d}x\mathrm{d}t,}_{\longrightarrow 0 \quad \text{for} \quad m \longrightarrow \infty \quad (\text{cf.} \quad (5.18)(b))}
\end{aligned}
$$

using Hölder with $1/r + 1/r^* = 1$. □

Remark. In practice, existence of the constants k_∞ and k_θ is no restrictive assumption. Bounding $\|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)}$ is basically assuming a finite maximum speed, which is a physically necessary assumption. Moreover, assuming a bound on $\|\nabla \cdot v\|_\theta$ in an applied sense is bounding the compressibility of the flow, which is again a reasonable assumption.

Remark. The operator A_t is defined on $L^l(\Omega)$ for $l \leq \frac{n}{n-1}$. This limitation of topologies compatible with the operator is necessary in order to ensure the continuous embedding of $BV(\Omega)$, which is the spatial domain in which we are searching for a minimizer, into the definition space of the operator.

For computational purposes, either model constraint is incorporated into the variational problem (5.4) by adding an additional proper, convex and lower semicontinuous penalty term T defined by

$$
T\big(c(\cdot,t), v(\cdot,t)\big) = \|m_i(c,v)(\cdot,t)\|^s_{L^r(\Omega)}, \quad r,s \geq 1, \quad t \in [0,T] \tag{5.20}
$$

for $i \in \{1,2\}$, such that we arrive at the unconstrained minimization problem

$$
\begin{aligned}
\min_{c,v} \int_0^T & D\big(A(c,t), u(t)\big) + \alpha R\big(c(\cdot,t)\big) + \beta S\big(v(\cdot,t)\big)\\
&+ \gamma T\big(c(\cdot,t), v(\cdot,t)\big)\mathrm{d}t,
\end{aligned} \tag{5.21}
$$

which can be solved by means of alternating minimization. The following lemma shows that solutions of the unconstrained problem (5.21) converge to solutions of the constrained problem (5.4) - (5.5) if the weighting parameter $\gamma \to \infty$.

Lemma 5.1: Convergence of solutions of the unconstrained minimization problems

Let

$$
\begin{aligned}
J_\gamma(c,v) := \int_0^T & D\big(A(c,t), u(t)\big) + \alpha R\big(c(\cdot,t)\big) + \beta S\big(v(\cdot,t)\big)\\
&+ \gamma T\big(c(\cdot,t), v(\cdot,t)\big)\mathrm{d}t
\end{aligned}
$$

for $\gamma > 0$ and

$$J_\infty(c,v) := \begin{cases} \displaystyle\int_0^T D\big(A(c,t), u(t)\big) + \alpha R\big(c(\cdot,t)\big) \\ \quad + \beta S\big(v(\cdot,t)\big)\mathrm{d}t, & \text{if } m_i(c,v) = 0 \\ \infty, & \text{else.} \end{cases}$$

It then holds that

1. The functionals J_γ are equicoercive in $L^{\hat{p}}\big(0,T;BV(\Omega)\big) \times \Big\{ v \in L^q\big(0,T;BV(\Omega)^n\big) \Big| \; \|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \le k_\infty, \; \|\nabla \cdot v\|_\theta \le k_\theta \Big\}$. We denote this space by Ξ in the following.

2. On bounded sets in the weak-* topology of the space $L^{\hat{p}}\big(0,T;BV(\Omega)\big) \times L^q\big(0,T;BV(\Omega)^n\big) := \Upsilon$, it holds that

$$J_\infty = \Gamma - \lim_{\gamma\to\infty} J_\gamma.$$

By Theorem 2.9, this is sufficient for convergence of the minima, i.e.,

$$\lim_{\gamma\to\infty} \inf_{(c,v)\in\Upsilon} J_\gamma(c,v) = \inf_{(c,v)\in\Upsilon} J_\infty(c,v).$$

By the strict convexity of J_∞ it follows that the sequence of minimizers (c_γ^*, v_γ^*) of J_γ converges to the minimizer (c^*, v^*) of J_∞.

Proof. The proof follows the argumentation of [35, Lemma 3.7], but is generalized to both motion models (instead of considering optical flow only).
Existence of minimizers of J_∞ is shown in the proof of Theorem 5.1. By the same arguments one can show existence of minimizers of J_γ for $\gamma \ge 0$.
J_γ are equicoercive if and only if there exists lower semicontinuous and coercive Ψ such that $J_\gamma \ge \Psi$. This holds as $J_\gamma \ge J_0$ and J_0 is coercive in Ξ for $\alpha, \beta > 0$ by the choice of the regularizers. To show Γ-convergence, we have to consider both conditions.

Let $(c,v) \in \Xi$ and (c_γ, v_γ) be a sequence converging to (c,v) in the weak-* topology. If (c,v) is admissable, i.e., $m_i(c,v) = 0$, then $J_\infty(c,v) = J_0(c,v)$. It then holds that

$$J_\infty(c,v) = J_0(c,v) \le \liminf_{\gamma\to\infty} J_0(c_\gamma, v_\gamma) \le \liminf_{\gamma\to\infty} J_\gamma(c_\gamma, v_\gamma),$$

by the lower semicontinuity of J_0. If (c,v) is not admissable, i.e., $m_i(c,v) \ne 0$, it follows by the lower semicontinuity of T that

$$0 < T(c,v) \le \liminf_{\gamma\to\infty} T(c_\gamma, v_\gamma),$$

which yields

$$J_\infty(c,v) = \infty = \liminf_{\gamma\to\infty} J_\gamma(c_\gamma, v_\gamma).$$

The first condition for Γ-convergence is thus fulfilled in both cases.
For the second condition, we choose the recovering sequence $(c_\gamma, v_\gamma) = (c, v)$. This results in

$$\lim_{\gamma\to\infty} J_\gamma(c_\gamma, v_\gamma) = \lim_{\gamma\to\infty} J_\gamma(c, v) = \left\{ \begin{array}{ll} \infty, & \text{if } \; m_i(c, v) \neq 0 \\ J_0(c, v), & \text{if } \; m_i(c, v) = 0 \end{array} \right\} = J_\infty(c, v).$$

□

5.2. Regularization properties of the joint approach

Consider the unconstrained formulation (5.21) with D and T norm discrepancies, i.e.,

$$\begin{aligned} \min_{c,v} \int_0^T & \|A(c,t) - u(t)\|^p_{L^p(\Omega)} + \alpha R\big(c(\cdot,t)\big) + \beta S\big(v(\cdot,t)\big) \\ & + \gamma \left\| m_i\big(c(\cdot,t), v(\cdot,t)\big) \right\|_{L^1(\Omega)} \mathrm{d}t, \end{aligned}$$

for $p \in \{1, 2\}$ and $i \in \{1, 2\}$. This describes a nonlinear inverse problem with forward operator $F : \mathrm{dom}(F) \subset U \to Y_1 \times Y_2$ with

$$\big(c(\cdot,t), v(\cdot,t)\big) \mapsto \Big(A\big(c(\cdot,t),t\big), m_i\big(c(\cdot,t), v(\cdot,t)\big)\Big). \tag{5.22}$$

Throughout this chapter, we considered $U = L^{\hat{p}}\big(0, T; BV(\Omega)\big) \times L^q\big(0, T; BV(\Omega)^n\big)$ for $\hat{p} = \min(2, p)$, $p > 1$ and $q > 1$. A general analysis of the regularization properties of this approach might be carried out based on the theory for Tikhonov-type regularization of nonlinear problems in Banach spaces [55, 130]. In this section, we briefly state the requirements which have to be fulfilled in general in order to obtain a regularization of a nonlinear problem. In a second step, we restrict our setting to a special case, where those assumptions are met.
We follow the formulation in [124] to analyze the setting. To ensure existence and stability of regularized solutions as well as convergence to a penalty-minimizing solution (for appropriate regularization parameters), the following assumptions have to be fulfilled [124, Theorem 3.22, Theorem 3.23 and Theorem 3.26].

1. The Banach spaces U and $Y_1 \times Y_2$ are associated with topologies weaker than the norm topology.
2. The norm functionals are sequentially lower semicontinuous with respect to the considered topologies on Y_1 and Y_2.
3. The functionals R and S are convex and sequentially lower semicontinuous with respect to the considered topology on U.
4. The regularization functional is proper on the domain of F.
5. The sublevel sets are pre-compact with respect to the topology on U.
6. The sublevel sets are sequentially closed with respect to the topology on U and the restriction of F to the sublevel sets is sequentially continuous with respect to the topologies on U and $Y_1 \times Y_2$.

Consider now our specific problem formulation. The first item can be fulfilled easily by equipping the Banach spaces U and $Y_1 \times Y_2$ with the weak-* and the weak topology, respectively. Consider $U = L^{\hat{p}}\big(0,T; BV(\Omega)\big) \times L^q\big(0,T; BV(\Omega)^n\big)$ for $\hat{p} = \min(2,p)$, $p > 1$ and $q > 1$. Therefore, U describes a reflexive Banach space and the weak and weak-* topology coincide on U.
The second item follows directly as the norm functional is convex and (lower semi-) continuous and thus weakly lower semicontinuous.
We assume the penalty functionals R and S to be proper, convex and weak-* lower semicontinuous in order to fulfill items three and four. As U is a reflexive space, it suffices if R and S are proper, convex and lower semicontinuous, which is our standard assumption throughout the chapter.
The fifth item demands weak-* pre-compact sublevel sets. If the sublevel sets are bounded in the weak-* topology, they are pre-compact if and only if every sequence has a weak-* convergent subsequence. Now in our case, we showed boundedness of the sublevel sets in the proof of Theorem 5.1. It followed from the constraint $\|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k_\infty$. Moreover, on a reflexive space every bounded sequence has a weakly convergent subsequence and thus a weak-* convergent subsequence.
For item six, we first note that weak-* closedness of the sublevel sets was also shown in the proof of Theorem 5.1. The crucial part for showing the regularization properties of our approach is thus to show that the restriction of F to the sublevel sets is sequentially continuous with respect to the topologies on U and $Y_1 \times Y_2$. We thus have to show that F is weak-*-to-weak continuous. In the first component of F, we have a bounded linear operator mapping to $Y_1 = L^2(0,T;Y)$ for a reflexive Banach space Y. Bounded linear operators are weak-to-weak continuous, which is sufficient in this case as we map between reflexive spaces. For the second component however, we need weak-*-to-weak continuity with respect to the tuple (c,v), i.e., for $c_n \stackrel{*}{\rightharpoonup} c$ and $v_n \stackrel{*}{\rightharpoonup} v$, we need that

$$m_i\left(c_n, v_n\right) \rightharpoonup m_i\left(c, v\right) \quad \text{for } \; n \to \infty.$$

We start by considering well-definedness of the motion model. We need to ensure existence of temporal and spatial derivatives of c and spatial derivatives of v by the choice of the space U. We thus have to restrict ourselves to a reflexive space ensuring existence of the mentioned derivatives.
Possible choices are, e.g., $U = W^{1,\hat{p}}\big(0,T; W^{1,s}(\Omega)\big) \times L^2\big(0,T; W^{1,q}(\Omega)^n\big)$ or even more specific the Hilbert space $U = H^1\big(0,T; H^1(\Omega)\big) \times L^2\big(0,T; H^1(\Omega)^n\big)$.
Considering the weak-*-to-weak continuity, we first note that from weak-* convergence of (c_n, v_n) it follows convergence of the motion models in a distributional sense (see the proof of Theorem 5.1), i.e.,

$$\int_0^T \int_\Omega \left[m_i(c_n,v_n) - m_i(c,v)\right] \varphi \,\mathrm{d}x\mathrm{d}t \longrightarrow 0 \quad \text{for } \; n \to \infty \; \text{ and } \; \varphi \in \mathcal{C}_0^\infty\left([0,T] \times \Omega\right).$$

Now, for $Y_2^* = \mathcal{C}_0^\infty\left([0,T] \times \Omega\right)$ this coincides with weak convergence, but this is a too restrictive assumption. However, the space of test functions $\mathcal{C}_0^\infty([0,T] \times \Omega)$ is dense in $L^p([0,T] \times \Omega)$ for $p < \infty$. For $\hat{r} = \hat{t}$, it holds that $L^{\hat{r}}([0,T] \times \Omega) = L^{\hat{r}}(0,T; L^{\hat{t}}(\Omega))$. We thus consider $Y_2^* = L^{\hat{r}}(0,T; L^{\hat{t}}(\Omega))$ for $1 < \hat{r}, \hat{t} < \infty$. Then, for arbitrary $f \in Y_2^*$ there exists a sequence $\varphi_k \in \mathcal{C}_0^\infty([0,T]\times\Omega)$ such that $\varphi_k \to f$ for $k\to\infty$ with respect to the norm in Y_2^*. It then holds that

$$\begin{aligned}
&\lim_{n\to\infty}\int_0^T\int_\Omega [m_i(c_n,v_n)-m_i(c,v)]\, f\mathrm{d}x\mathrm{d}t\\
&=\lim_{n\to\infty}\int_0^T\int_\Omega [m_i(c_n,v_n)-m_i(c,v)]\,(f-\varphi_k)\,\mathrm{d}x\mathrm{d}t\\
&\quad+\int_0^T\int_\Omega [m_i(c_n,v_n)-m_i(c,v)]\,\varphi_k\,\mathrm{d}x\mathrm{d}t,
\end{aligned}$$

where the second term converges to zero as φ_k is a test function. For the first term, we observe

$$\begin{aligned}
&\lim_{n\to\infty}\left|\int_0^T\int_\Omega [m_i(c_n,v_n)-m_i(c,v)]\,(f-\varphi_k)\,\mathrm{d}x\mathrm{d}t\right|\\
&\leq \lim_{n\to\infty}\|m_i(c_n,v_n)-m_i(c,v)\|_{Y_2}\,\|f-\varphi_k\|_{Y_2^*}\,.
\end{aligned}$$

The norm of the first term is uniformly bounded if we assume a bound on the spatial derivative of c, i.e., $\|\nabla c\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k'_\infty$ additionally to $\|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)} \leq k_\infty$. Moreover, for each $\varepsilon>0$ there exists a $k\in\mathbb{N}$ such that $\|f-\varphi_k\|_{Y_2^*}<\varepsilon$.
We thus conclude that the operator $F:\mathrm{dom}(F)\to L^2(0,T;Y)\times L^r\big(0,T;L^r(\Omega)\big)$ for

$$\begin{aligned}
\mathrm{dom}(F) =& H^1\big(0,T;H^1(\Omega)\big)\times L^2\big(0,T;H^1(\Omega)^n\big)\cap\\
&\Big\{(c,v)\,|\,\|\nabla c\|_{L^\infty(0,T;L^\infty(\Omega)^n)}\leq k'_\infty \wedge \|v\|_{L^\infty(0,T;L^\infty(\Omega)^n)}\leq k_\infty\Big\},
\end{aligned}$$

describes a weak-*-to-weak continuous operator. A more general investigation is left for future work.

5.3. Alternating minimization

In this section, we consider the unconstrained minimization problem

$$\begin{aligned}
\min_{c,v} J(c,v) = \int_0^T & D\big(A(c,t),u(t)\big)+\alpha R\big(c(\cdot,t)\big)+\beta S\big(v(\cdot,t)\big)\\
&+\gamma T\big(c(\cdot,t),v(\cdot,t)\big)\mathrm{d}t.
\end{aligned}$$

The problem is solved by means of alternating minimization. This means that instead of solving for the tuple (c,v) directly, one solves for c and v alternately, i.e.,

$$\min_c \int_0^T D\big(A(c,t),u(t)\big)+\alpha R\big(c(\cdot,t)\big)+\gamma T\big(c(\cdot,t),v(\cdot,t)\big)\mathrm{d}t, \tag{5.23}$$

and

$$\min_v \int_0^T \beta S\big(v(\cdot,t)\big)+\gamma T\big(c(\cdot,t),v(\cdot,t)\big)\mathrm{d}t, \tag{5.24}$$

are solved alternately holding the other variable fix. We call (5.23) the *image reconstruction subproblem* and (5.24) the *motion estimation subproblem*.

We thus solve two convex problems which are formulated with help of linear operators. The alternative of solving the joint problem directly is difficult as the problem has a nonlinearity due to the nonlinear motion model and is non-convex, which might lead to local minima. Standard methods as ADMM or split Bregman cannot be applied due to the nonlinear term.

Remark. Convergence (to a global minimizer) of the alternating minimization scheme cannot be shown in general. Denoting the solutions of the subproblems in the n-th iteration by c_n and v_n, we see that $J(c_n, v_n) \geq J(c_{n+1}, v_n) \geq J(c_{n+1}, v_{n+1})$, i.e., the sequence is decreasing and bounded from below for admissable initial values (c_0, v_0). Therefore, the sequence converges to some $b \geq \inf_{(c,v)} J(c, v)$. In practice, this convergence is usually very fast. For a strongly bi-convex functional J, it is clear that we converge to a global minimum. Under less restrictive assumptions, this is not the case in general. Convergence to a first order stationary point can be shown, e.g., under the assumption that J satisfies the Kurdyka-Łojasiewicz property, is bi-convex, has a Lipschitz continuous gradient on any bounded subsets of the domain and moreover, the sequence (c_n, v_n) is bounded [8]. In [16], convergence properties are analyzed for a functional with the part depending on both variables being differentiable.

In our case, we can only expect convergence to a local minimum and thus our result depends largely on the initial guess. However, in practice the algorithm yields convincing results.

In the remainder of this section, we consider the algorithmic solutions to the subproblems.

We can see directly, that the motion estimation subproblem coincides with the problems formulated in Chapter 4. We thus solve the problem with PDHG for either optical flow or mass conservation, using a multi-scale and warping or a multi-scale only approach.

For the image reconstruction subproblem, the formulation is slightly more complex as the ones we have encountered in Section 3.3. We thus briefly derive the scheme here, limiting ourselves to non-negative fused lasso regularization and an L^1 data fitting term in the following. As we have seen in Chapter 4, warping can be applied in the optical flow case. We therefore implement the nonlinearized version of the gray-value constancy assumption according to (4.1), i.e., in the semi-discretized setting with respect to time step h_t we use

$$\begin{aligned} T\big(c\,(\cdot,t)\,,v\,(\cdot,t)\,\big) &= \big\|c\big(\,\cdot+h_t v\,(\cdot,t)\,,t+h_t\big) - c(\cdot,t)\big\|_{L^1(\Omega)} \\ &= \|\mathcal{W}c\,(t)\|_{L^1(\Omega)}\,. \end{aligned}$$

with

$$(\mathcal{W}c)\,(t) := -c\,(\cdot,t) + W^{t,h_t}c\,(\cdot,t+h_t)\,,$$

where W^{t,h_t} maps a space-dependent image to a space-dependent image performing the reverse warping of the input image with respect to the displacement field estimate $v(\cdot,t)$ and the time step h_t. Similar to the motion estimation subproblem, reverse warping (iterating over the destination image) is performed in order to cover all pixels of the destination image. Applying primal-dual splitting to this setting and using the same notations as in the motion estimation subproblem leads to

$$f_1(c) = \int_0^T \alpha_1 \|c(\cdot,t)\|_{L^1(\Omega)} + \mathbb{I}_{\{c(\cdot,t)\geq 0\}} \,\mathrm{d}t$$
$$f_2(Cc) = \int_0^T \|A(c,t) - u(t)\|_{L^1\left(\mathbb{R}^d\right)} + \alpha_2 \|\nabla c(\cdot,t)\|_{L^1(\Omega)^n} + \gamma \|(\mathcal{W}c)(t)\|_{L^1(\Omega)} \,\mathrm{d}t$$

with the linear operator $C = \begin{pmatrix} A, \nabla, \mathcal{W} \end{pmatrix}^T$ and the corresponding adjoint operator $C^* = \begin{pmatrix} A^*, & -\mathrm{div}, & \mathcal{W}^* \end{pmatrix}$. The corresponding Fenchel-conjugate is given by

$$f_2^*(y_1, y_2, y_3) = \int_0^T \mathbb{I}_{\left\{\|y_1\|_\infty \leq 1\right\}} + \langle y_1, u\rangle + \mathbb{I}_{\left\{\|y_2\|_{3,\infty} \leq \alpha_2\right\}} + \mathbb{I}_{\left\{\|y_3\|_\infty \leq \gamma\right\}} \mathrm{d}t. \tag{5.25}$$

The mass conservation constraint

$$m_2(c,v) = \frac{\partial}{\partial t} c + \nabla \cdot (cv) = 0.$$

is not linearized. We therefore use the constraint directly inside our penalty term, i.e.

$$T\big(c(\cdot,t), v(\cdot,t)\big) = \left\|\frac{\partial}{\partial t} c(\cdot,t) + \nabla \cdot \Big(c(\cdot,t)\, v(\cdot,t)\Big)\right\|_{L^1(\Omega)}$$
$$= \left\|\frac{\partial}{\partial t} c(\cdot,t) + c(\cdot,t)\, \nabla \cdot \Big(v(\cdot,t)\Big) + v(\cdot,t) \cdot \nabla c(\cdot,t)\right\|_{L^1(\Omega)}.$$

Applying again the notation from the motion estimation subproblem leads to

$$f_1(c) = \int_0^T \alpha_1 \|c(\cdot,t)\|_{L^1(\Omega)} + \mathbb{I}_{\{c(\cdot,t)\geq 0\}} \,\mathrm{d}t$$
$$f_2(Cc) = \int_0^T \|A(c,t) - u(t)\|_{L^1\left(\mathbb{R}^d\right)} + \alpha_2 \|\nabla c(\cdot,t)\|_{L^1(\Omega)^n}$$
$$+ \gamma \left\|\frac{\partial}{\partial t} c(\cdot,t) + c(\cdot,t)\, \nabla \cdot \Big(v(\cdot,t)\Big) + v(\cdot,t) \cdot \nabla c(\cdot,t)\right\|_{L^1(\Omega)} \mathrm{d}t,$$

where

$$C = \begin{pmatrix} A \\ \nabla \\ \frac{\partial}{\partial t} + v_1 \frac{\partial}{\partial x} + v_2 \frac{\partial}{\partial y} + v_3 \frac{\partial}{\partial z} + (\nabla \cdot v)\, I \end{pmatrix},$$

and

$$C^* = \begin{pmatrix} A^*, & -\mathrm{div}, & -\left(\frac{\partial}{\partial t} + v_1 \frac{\partial}{\partial x} + v_2 \frac{\partial}{\partial y} + v_3 \frac{\partial}{\partial z} - (\nabla \cdot v)\, I\right) \end{pmatrix}.$$

The Fenchel-conjugate corresponding to f_2 is the same as in (5.25).
Inserting the functions f_1 and f_2^* into the primal-dual framework in Algorithm 1 yields the PDHG algorithm for the image reconstruction subproblem. It is further possible to apply splitting of the dual variable corresponding to the data term and apply SPDHG algorithm. The alternating scheme for joint image reconstruction and motion estimation is illustrated by Algorithm 6.

Algorithm 6 Alternating image reconstruction and motion estimation

Input: forward operator A, data u, parameters $\alpha, \beta > 0$, initial image sequence c^0, parameters $\gamma, \delta > 0$, initial motion field guess $v = v^0$, number of scales Z, motion model m_i
for k=1,2,3,... **do** $\{$*% outer iterations = joint iterations*$\}$
  $c^k = \text{PDHG for image reconstruction subproblem}\left(A, u, c^{k-1}, \alpha, \beta, \gamma, v^{k-1}, m_i\right)$
  $v^k = \text{PDHG for motion estimation}\left(c^k, \gamma, \delta, v^{k-1}, Z, m_i\right)$
end for

5.4. Summary and discussion

Joint motion estimation and image reconstruction is expected to increase the quality of both motion estimates and reconstructed image sequences. In this chapter, we fixed the problem setting and formulated Theorem 5.1. This result forms the main contribution of this chapter and has been published before in [27]. The theorem follows the idea of [35, 50] but extends the setting from 2D plus time to 3D plus time image sequences and considers time-dependent linear forward operators. The existence of a minimizer to the joint image reconstruction and motion estimation task is proven. Therefore, the motion model is incorporated into the variational problem formulation as a hard constraint that holds almost everywhere. Afterwards, we incorporated the motion model into the variational scheme as an additional penalty term and showed convergence of such unconstrained solutions to solutions of the originally constrained problem. We solved the corresponding optimization problem by alternating minimization using the algorithms derived in the previous chapters for image reconstruction and motion estimation, respectively. Moreover, we briefly considered under which additional assumptions the proposed joint approach can be seen as a regularization of a nonlinear inverse problem. However, we only considered quite restrictive assumptions in order to guarantee the weak-*-to-weak continuity of the nonlinear forward operator. A more general discussion in a more complex and less limiting framework is an open question left for future work.

6. Experiments on Synthetic and Measured Data

This chapter is dedicated to extensive numerical tests of the proposed joint approach, in particular in comparison to standard static reconstruction methods. Our goal is to show the applicability of the proposed joint approach to reconstruction of dynamic MPI images. All implementations are coded in MATLAB. This chapter is structured as follows. We start with experiments on simulated data for which we have a ground truth solution available in Section 6.1. This ground truth solution exists not only for the reconstructed images but also for the displacement fields. We can thus compare our results to the ground truth in terms of different quality measures. Afterwards, we consider two different sets of measured data. First, we use in-vitro data measuring a rotation phantom which fulfills the OF constraint in Section 6.2. As this phantom fulfills the OF constraint, it also fulfills the MC constraint. Second, we reconstruct in-vivo data covering the cardiovascular system of a mouse in Section 6.3. Images from this dataset fulfill the MC constraint but not the OF constraint. The experiments of these three section have mainly been carried out in preparation of [27] and most results have been first published therein. We extend our considerations on simulated data by taking into account quality measures for the motion estimates, whereas in [27] we considered quality measures for the reconstructed images only. Moreover, we analyze the evolution of results of the joint approach over joint (outer) iterations and add an investigation of runtime and bottlenecks of the implemented algorithms. Considering the rotation phantom data, we additionally present coefficients of variation for the 1 Hz and 7 Hz dataset and we state the radii of reconstructed trajectories as further quality measure. We conclude this chapter by Section 6.4, where we consider subframe reconstruction. In this setting, we do not use data from a complete frame for image reconstruction, but only parts of it in order to enhance the temporal resolution of the resulting image sequences. We show that the proposed joint approach provides this option on simulated and measured data.

6.1. Joint reconstruction versus a two-step scheme - experiments on 3D simulated data

In this section, we reconstruct from synthetic data in order to have a ground truth available. The MPI scanner is modeled based on the pre-clinical MPI scanner (Bruker Biospin, Ettlingen) at the University Medical Center Hamburg-Eppendorf, and a detailed list of parameter choices can be found in Section A.1.1. Our phantom resembles the rotation phantom presented in [64] and has a temporal resolution of 26928 states per frame. It consists of a ball moving along a circular trajectory, thus fulfilling the OF constraint. We simulate data on a grid consisting of $40 \times 40 \times 40$ voxels and reconstruct on a smaller grid of $20 \times 20 \times 20$ voxels. Further, we consider three different noise levels: a noise-free scenario, a medium noise scenario and a high noise scenario. Noisy data is obtained by adding Gaussian noise with mean value zero and standard deviation 0%, 50% and 100% of the maximum absolute signal to the data in the time domain. Afterwards, the data is Fourier transformed and pre-processed in the frequency domain. The pre-processing includes limiting the maximum mixing order to 15, the maximum mixing factors to 10 and concatenating real and imaginary part of the data. The obtained data is as a result corrupted by random noise with standard deviation of 0%, 2.7% and 5.3% of the maximum absolute signal. The noise-free scenario is contained in order to analyze the influence of noisy data on the algorithm but is not relevant in practice. Depending on the experimental setup, both the medium and high noise scenario might resemble a realistic scenario. However, our simulation setup uses a simple forward model based on the Langevin magnetization model and neglects further noise sources as for example background signals. Moreover, in contrast to a forward operator in practice, which is typically measured and thus noisy, the simulated forward operator is not corrupted by noise. The reconstruction task is thus performed under ideal conditions compared to the reconstruction of measured data even in the case of noisy simulated data.

In the proposed joint reconstruction setting, we consider four different algorithms. All of them apply non-negative fused lasso regularization on the tracer concentration (cf. (3.9)) and total variation regularization on the motion estimates (cf. (4.8)). As data discrepancy terms, we use L^1- and L^2-norms and the motion model is either the OF constraint or the MC constraint. The resulting algorithms are listed in Table 6.1 and denoted by L1OF, L2OF, L1MC and L2MC.

Table 6.1.: Overview of joint reconstruction schemes consisting of an image reconstruction algorithm and a motion estimation algorithm. We state the incorporated data discrepancy term as well as the motion model used.

Abbr.	Data disc. term	Image rec. alg.	Motion model	Motion estimation alg.
L1OF	L^1-norm	PDHG	Optical flow	Multiscale & warping PDHG
L2OF	L^2-norm	PDHG	Optical flow	Multiscale & warping PDHG
L1MC	L^1-norm	PDHG	Mass conservation	Multiscale PDHG
L2MC	L^2-norm	PDHG	Mass conservation	Multiscale PDHG

All approaches are tested on 10 frames simultaneously. The quality of the reconstructed image sequences is assessed by SSIM, PSNR and Mean Squared Error (MSE). The displacement fields are compared to the ground truth in terms of L^2-norm difference and AAE.

First of all, we compare a joint reconstruction approach to a two-step scheme, where image reconstruction is performed first and motion estimation second. This also allows us to investigate how many joint iterations are needed and serves as a proof of concept for the joint approach.

In Figure 6.1, we illustrate the evolution of the reconstructed image frames during the joint iterations. The visual perception confirms the increase of the SSIM during the first three iterations. The effect of applying more iterations is negligible. These results allow

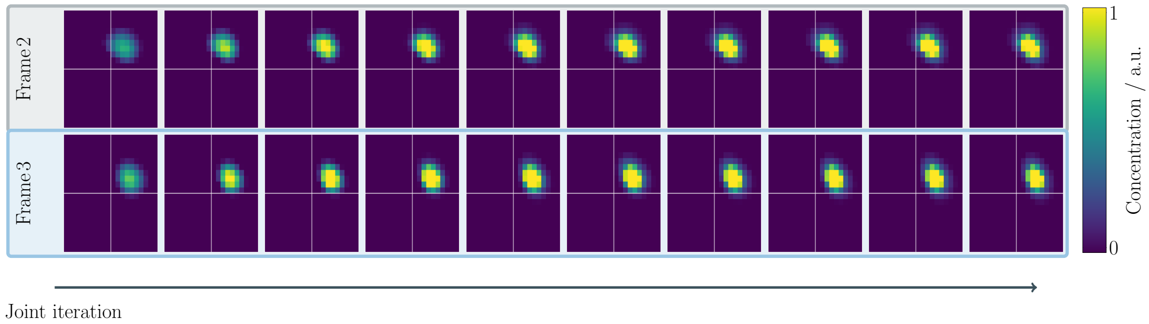
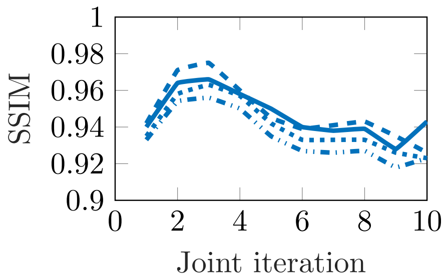
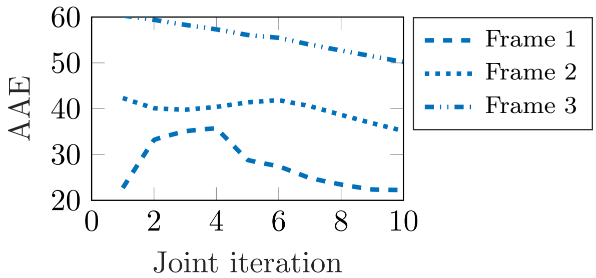


Figure 6.1.: Reconstructed images of the second and third frame are depicted in the upper and bottom row, respectively. The columns show the intermediate results after one to ten joint iterations. An increase in the visually perceived quality of the reconstructed images is obtained for the first three to four iterations, for a higher number of iterations the differences are visually imperceptible.



(a) SSIM value.



(b) AAE value.

Figure 6.2.: Applying the L1OF algorithm for image reconstruction and motion estimation to the first four frames yields four consecutive images and motion estimates for the three intervals in between. The image sequence yields the SSIM value in (a) and motion estimates obtain the AAE in (b).

for early stopping of the joint algorithm after three iterations in order to obtain the highest possible image quality.

In Figure 6.2, we depict the SSIM and AAE values for the first ten joint iterations for the Joint reconstruction algorithm using an L^1-data discrepancy term and the optical flow motion model (L1OF) algorithm. The SSIM increases first, but starts decreasing after three iterations. The AAE indicates that the motion estimation is not accurate, the decrease in the SSIM might thus be caused by overfitting to a non-accurate motion estimate. Clearly, early-stopping is a reasonable regularization method to avoid this behavior.

Let us now consider two-step schemes in comparison. A two-step scheme can be based on frame-by-frame image reconstruction by the Kaczmarz algorithm using standard Tikhonov L^2- regularization combined with a positivity constraint [48], which corresponds to one of the most popular MPI reconstruction techniques [135]. Additionally, we consider a two-step scheme based on reconstructed images by SPDHG algorithm solving a formulation with non-negative FL regularization. Both algorithms were introduced and tested comprehensively in Section 3.3. We combine both schemes with the multiscale and warping PDHG-OF, which is the natural choice as the phantom fulfills the OF constraint.

Using the Kaczmarz algorithm for image reconstruction yields image sequences with significantly lower SSIM, reaching a maximum value of approximately 0.66. The maximum concentration value is strongly underestimated, better results in terms of that value reach a SSIM of only approximately 0.3. The motion estimation task is then performed on that image sequence as input data and reaches an AAE of 4/44/80 degrees for the first three frames, respectively. Although this value is very good for the first frame, the other frames are worse than results by the joint approach. Moreover, the L^2-error is significantly higher (0.88 compared to 0.77 after three joint iterations). When reconstructing images by SPDHG, the image quality resembles the joint approach. The motion estimation task reaches an AAE of 39/40/58 for the first three frames and an L^2-error of 0.93, resulting in an overall slightly worse performance compared to the joint algorithm.

We now consider the quality of the image sequences obtained by the different approaches in more detail. We perform an extensive parameter search and choose the ones obtaining the highest SSIM values. The resulting reconstruction parameters are given in Section A.1.2. Table 6.2 states the mean SSIM, PSNR and MSE values for all algorithmic approaches and the three different noise scenarios. For illustration, reconstructed images for the second time step and the medium noise level are depicted in Figure 6.3. For all three noise scenarios, the Kaczmarz method yields by far the worst quality images. This is clear from the visual impression but also underlined by all three quality measures. The images reconstructed by the Kaczmarz algorithm suffer from severe motion and noise artifacts. The SPDHG reconstructed images benefit from the FL regularization, which handles the noise much better due to the sparsity enforcing L^1-norm and the TV term. It reaches higher SSIM and PSNR values than the joint approaches using an L^2-data discrepancy term for the noise-free and medium noise scenario. For the high noise scenario, all joint algorithms outperform the two static image reconstruction algorithms. Best results are obtained using an L^1-data discrepancy term, that is able to handle outliers in the data more robustly.

Table 6.2.: Comparison of the different algorithms in terms of SSIM, PSNR and MSE of the reconstructed image sequences for three different noise levels and simulated data. The table was first published in [27].

Noise level	Noise-free			Medium			High		
	SSIM	PSNR	MSE $\cdot 10^{-4}$	SSIM	PSNR	MSE $\cdot 10^{-4}$	SSIM	PSNR	MSE $\cdot 10^{-4}$
Kaczmarz	0.657	25.31	30	0.654	23.18	48	0.651	22.14	61
SPDHG	0.955	27.12	19	0.953	26.93	20	0.946	26.52	23
L1OF	**0.962**	27.79	**17**	**0.960**	27.56	**18**	**0.955**	**27.62**	**18**
L2OF	0.947	26.61	22	0.946	26.51	22	0.950	27.02	20
L1MC	**0.963**	**27.85**	**17**	**0.961**	**27.60**	**18**	**0.956**	27.57	**18**
L2MC	0.947	26.60	22	0.946	26.49	23	0.950	27.00	20

From these experiments, we could question the fact that Kaczmarz method defines the state-of-the-art reconstruction method in MPI. However, in a standard static setting measurements over multiple frames can be averaged in order to increase the SNR of the data. As MPI has a fast acquisition time, it does not pose a problem to measure, e.g., 100 DF cycles in order to average the obtained signal. Having a reconstruction algorithm

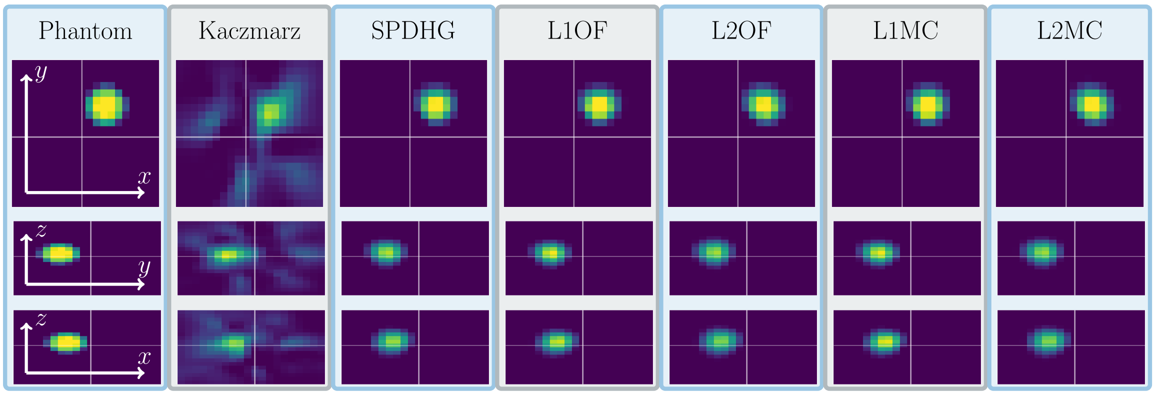


Figure 6.3.: Reconstructed images of the simulated data for the second time step and the medium noise level. The upper row depicts averaged intensity projections onto the x-y-plane, the middle row projects onto the y-z-plane and the bottom row onto the x-z-plane. Each column corresponds to one reconstruction algorithm. All images are plotted with the same colorbar viridis and windowed based on the full dynamic range of the phantom. The joint approaches as well as SPDHG yield noise-free image sequences which slightly underestimate the maximum concentration values, but recover the phantom well. The image sequence reconstructed by Kaczmarz method underestimates the concentration stronger and suffers from background artifacts which make it difficult to distinguish between phantom and artifacts especially in the y-z- and x-z-planes. The figure was first published in [27].

that handles noise effectively is then not as important as having a fast reconstruction algorithm, at the best of times allowing for online reconstruction. However, if we apply frame averaging in the dynamic case, this results in severe motion artifacts as can be seen in Figure 6.4. For averaging of very few frames (five in this example) the SNR in the data is barely improved and motion artifacts occur already, but for averaging of 200 frames the motion artifacts make it impossible to locate the object. This behavior was observed by [64] as well.

Still, a great advantage of the Kaczmarz method is its fast reconstruction time. SPDHG can challenge its runtime in a static setting, as we explored extensively in [146]. However, it is necessary to tune the step size parameters accurately and find a reasonable early stopping time in order to be competitive. Most computational costs (about 94%) arise from the matrix-vector product involving the forward operator when performing a data step update. The joint approaches, however, are by far slower than the Kaczmarz method. Let us first consider the OF constrained case. As currently implemented, one joint iteration of L1OF consisting of 10^2 iterations within the image reconstruction task and 10^3 iterations within the motion estimation task, needs approximately seven times as much computation time compared to Kacmarz method (three iterations). Thereby, the image reconstruction task covers more than 95% of the run time. Within the image reconstruction algorithm, 30% of the computational cost is allotted to the matrix-vector products again. The main cost is attributed to inversion of the displacement mapping, which makes use of a scattered interpolant. A computationally more efficient implementation of this step would imply huge savings. As the joint algorithms based on the MC constraint do not perform the inversion and thus do not use a scattered interpolant, they are significantly faster. Still, one joint iteration has almost double the cost of the Kaczmarz method. Within one joint iteration, about 80% of the run time are allotted to the image reconstruc-

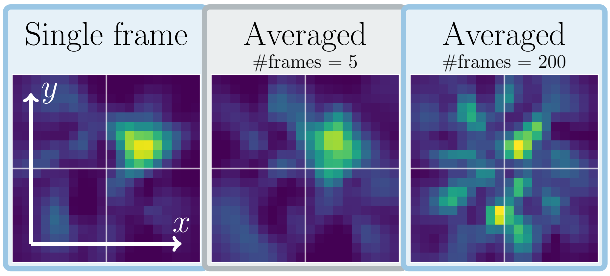


Figure 6.4.: Averaged intensity projections of the reconstructed images of the rotation phantom for the medium noise level. The reconstructions are obtained via Kaczmarz method using single frame measurements (left column) as well as averaged measurements over multiple frames (middle and right column). The best regularization parameter (obtaining the highest SSIM value) λ_{single} weighting the Tikhonov regularization term was determined for the single frame reconstruction and then adjusted to the amount of averaging (using $\lambda_{\text{single}}/\sqrt{\#\text{frames}}$) in order to stay proportional to the expected noise level. The amount of frame averaging is increased from the left to the right column. Note that the reconstructions are windowed differently, the maximum concentration is 5/3/1.5 from left to right in absolute numbers. Although averaged measurements obtain a higher signal-to-noise ratio, the quality of the reconstructed images does not increase from left to right due to the dynamic nature of the phantom. The figure was first published in [27] but is slighty altered.

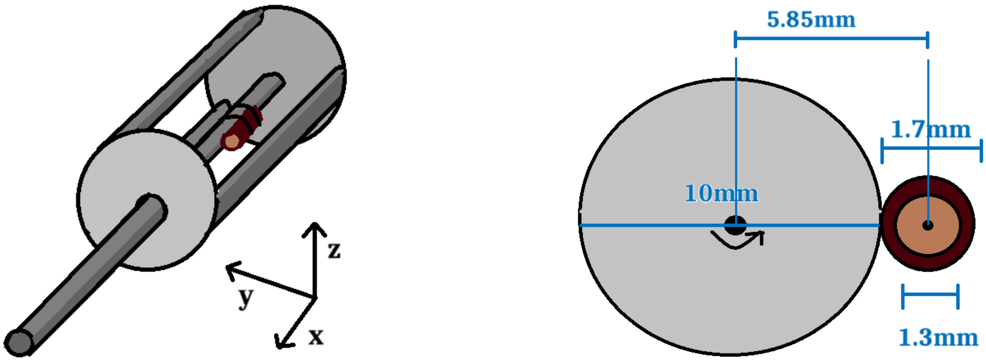
tion task, whereas the motion estimation task is again comparably cheap. The main cost factor during the image reconstruction task is again the matrix-vector product involving the forward operator. The bottleneck in the motion estimation algorithm PDHG-OF is given by the differential operators. By porting the motion estimation task from MATLAB to C++ the author of [50] was able to produce a speed-up of factor approximately 50.

6.2. In-vitro rotation phantom fulfilling the optical flow constraint

The rotation phantom data used in this section were first published in [64] in 2017. We thank the group of Prof. Dr. Tobias Knopp of the University Medical Center Hamburg Eppendorf and the Technological University of Hamburg for providing it.

The measurements were taken by a pre-clinical MPI scanner (Bruker Biospin, Ettlingen) using sinusoidal excitation frequencies creating 3D Lissajous trajectories. The drive-field excitation frequencies are $f_x = 2.5\,\text{MHz}/102$, $f_y = 2.5\,\text{MHz}/96$ and $f_z = 2.5\,\text{MHz}/99$ with an amplitude of $14\,\text{mT}/\mu_0$, resulting in a repetition time of $21.54\,\text{ms}$. The selection field gradient strength is $0.75\,\text{T/m}/\mu_0$ in x- and y-direction and $1.5\,\text{T}/(\text{m}\mu_0)$ in z-direction. The delta sample used for system matrix acquisition was of size $2 \times 2 \times 1\ \text{mm}^3$ and covers $25 \times 25 \times 25$ positions.

For construction of the phantom, two round disks of the scanning bores' diameter are connected by three rods. An additional rod is placed in the center. On this rod, a glass capillary with an inner diameter of $1.3\,\text{mm}$, an outer diameter of $1.7\,\text{mm}$ and filled with $20\,\mu\text{L}$ diluted ferucarbotran with factor $1/10$ is fixed. The sample itself thus has a size of approximately $1.3\,\text{mm}^2 \times 9.8\,\text{mm}$. The central rod with a diameter of $10\,\text{mm}$ can be attached to a screwdriver and rotated during measurements. Rotation of the central rod thus results in a circular path of the tracer material with a diameter of approximately 11.7 mm. Measurements are available for the different rotation speeds, the respective motion frequencies are 1 Hz, 3 Hz and 7 Hz. A schematic illustration of the phantom is depicted in Figure 6.5. A more detailed description of the phantom as well as reconstructions for comparison can be found in [64].



(a) Setup of the rotation phantom.

(b) Distances of capillary and rods.

Figure 6.5.: Schematic illustration of the rotation phantom.

We use a standard procedure for data pre-processing, applying an SNR threshold of five, and cutting off frequencies lower than 80 kHz. As for the simulated data, we concatenate real and imaginary parts in order to have a convenient format for the primal-dual image reconstruction algorithm. Moreover, we apply row normalization as introduced in [93] as a weighting approach.

Both motion models are applicable such that we compare all four joint approaches on the data. In Figure 6.6, we depict exemplary reconstructed images of the 3 Hz dataset for the four joint approaches as well as for the static approaches by Kaczmarz method and SPDHG. Each row shows projections onto the y-z-plane for one frame in time. Projections onto the y-x- and z-x-plane are depicted in Figure 6.7.

Figure 6.6 shows that variational approaches applying FL regularization yield image sequences that are significantly less noisy than images reconstructed by the Kaczmarz method. In the y-z-plane, we would ideally have reconstructed phantoms of size one to two voxels in each direction. The variational approaches are close to this, whereas the dot shape phantom is blurry when reconstructed by the Kaczmarz method. The differences between L^1- and L^2-data term are negligible, but reconstructed images by an L^1-data term are slightly sparser. Frame-by-frame SPDHG performs well in terms of size of the object and background noise, however, the algorithm faces severe problems in achieving comparable mass in each frame. This impression is confirmed by the coefficient of variation, i.e., standard deviation divided by mean value, of the reconstructed mass in each frame over the first 30 frames for each reconstruction approach in Table 6.3. Where SPDHG faces severe issues, the other approaches perform well.

Considering Figure 6.7 and projections onto the y-x- and z-x-plane, we aim at a phantom of length five to seven voxels in x direction. The proposed joint schemes reconstruct a phantom of reasonable size, more particularly of approximately seven voxels in length and two voxels in width. For the Kaczmarz method and SPDHG, the phantom seems to

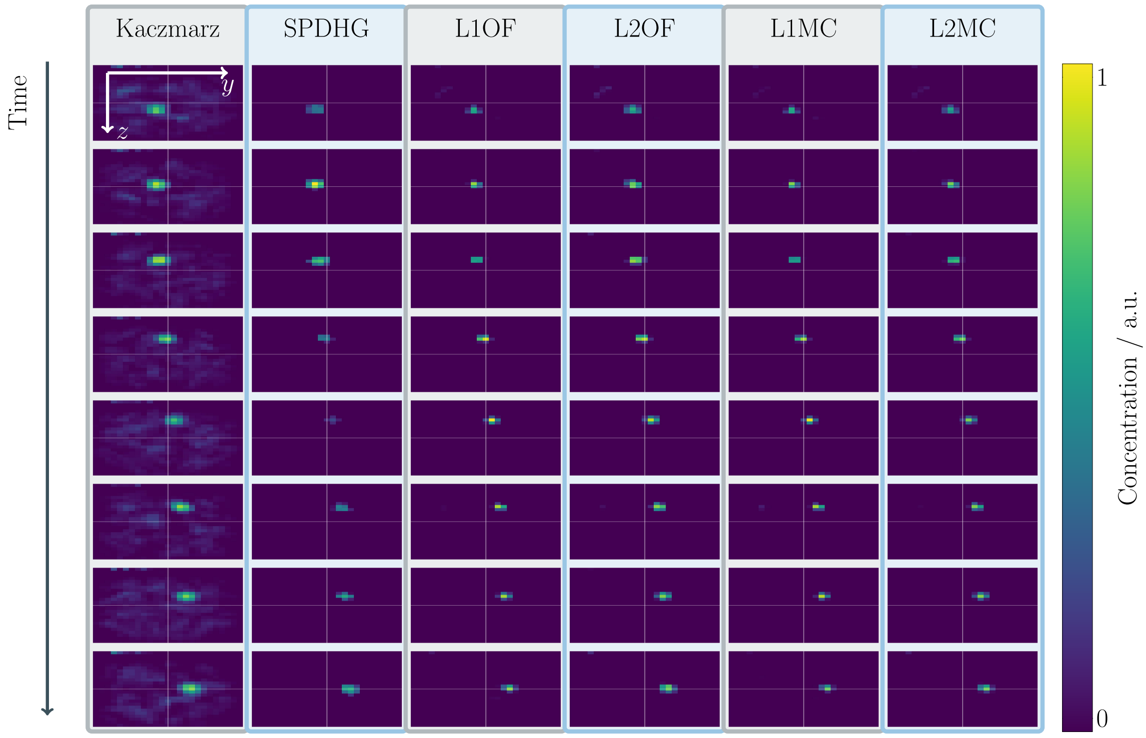


Figure 6.6.: Reconstructed image sequences for the 3 Hz rotation dataset. Each column corresponds to one algorithm, each row represents one frame in time. We display projections onto the y-z-plane in this figure. The joint approaches yield sparser reconstructed image sequences compared to the Kaczmarz method. Moreover, they feature an empty background whereas the images reconstructed by Kaczmarz method suffer from background noise. The figure was first published in [27].

Table 6.3.: Coefficients of variation of the total reconstructed mass in each frame over the first 30 frames of the image sequences for the different rotation phantom datasets.

Speed	Kaczmarz	SPDHG	L1OF	L2OF	L1MC	L2MC
1 Hz	0.110	0.645	0.140	0.096	0.070	0.128
3 Hz	0.149	0.556	0.116	0.141	0.110	0.129
7 Hz	0.096	0.159	0.173	0.172	0.175	0.243

vary strongly in size. In time frames four to six, it appears significantly shorter than in other frames. Moreover, the reconstructions suffer from noise artifacts and varying mass, respectively.

All reconstruction parameters are given in Table 6.4. They were chosen based on visual inspection out of a wide range of tested parameters. The joint algorithm proposed here is used in combination with early stopping in order to limit the computation time. However, the reconstructed image sequences are close to convergence at that point at least based on visual perception we cannot observe any changes. More precisely, the early stopping iteration is given for the image reconstruction subtask in the inner loop of the alternating algorithm, cf. Section 5.3 and Algorithm 6. The alternations are performed three times.

Having assessed the image quality, we now investigate the impact of the proposed joint approach to the motion estimation task. Knowing that the particles follow a circular trajectory with known width and approximately known position, we decide to analyze the trajectories for comparison of the motion estimates. In a first step, we obtain the best

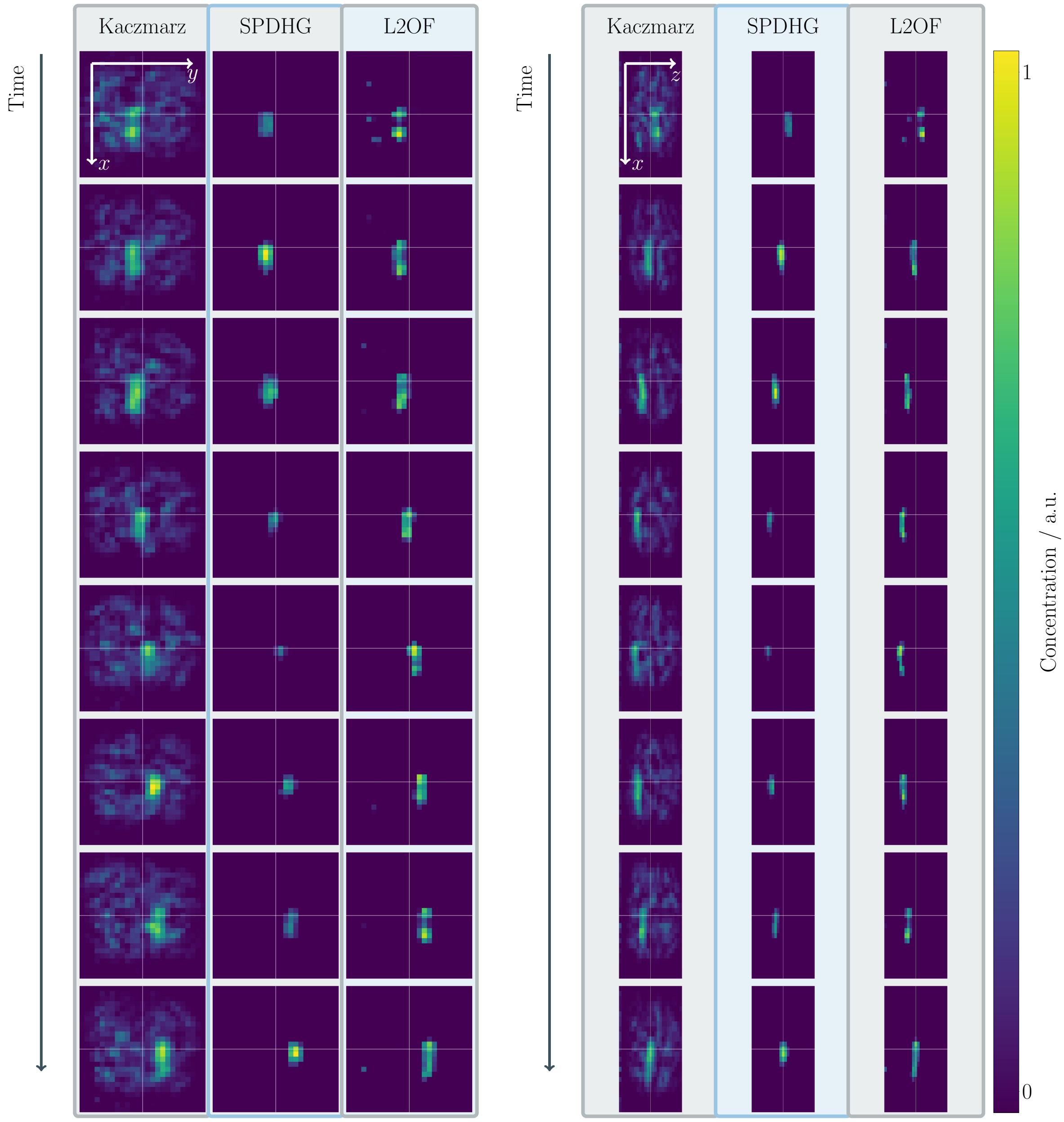


Figure 6.7.: Reconstructed image sequence of the 3 Hz rotation phantom for the Kaczmarz method, SPDHG and the L2OF algorithm. This figure displays projections onto the y-x- and z-x-planes for the first eight time frames. The Kaczmarz method yields noise-corrupted reconstructed images, whereas variational approaches result in clean image sequences. The phantom reconstructed by SPDHG varies strongly in size and underestimates the phantom. The joint L2OF algorithm captures the correct size of the phantom. The figure was first published in [27].

Table 6.4.: The parameters λ, α, β and the early stopping iteration index k used for the reconstruction of the 3 Hz rotation phantom data. The parameters $\gamma = 100$ (motion model penalty) and $\delta = 10^{-1}$ (TV regularization on the motion field) are the same for all joint approaches.

Algorithm	Early stopping	λ	α (L^1 on image)	β (TV on image)
Kaczmarz	$k = 10$	5.62		
SPDHG	$k = 10^3$		$3.0 \cdot 10^{-1}$	$5.0 \cdot 10^{-1}$
L1OF	$k = 10^2$		$1.0 \cdot 10^{-1}$	$6.0 \cdot 10^{-1}$
L2OF	$k = 10^2$		$1.0 \cdot 10^{+2}$	$2.5 \cdot 10^{-1}$
L1MC	$k = 10^2$		$5.0 \cdot 10^{-1}$	$7.0 \cdot 10^{-1}$
L2MC	$k = 10^2$		$2.5 \cdot 10^{-1}$	$2.5 \cdot 10^{-1}$

possible trajectories from the corresponding reconstructed image sequences manually by following the maximum concentration along the time frames. The resulting trajectories are smoothed by a fourth order polynomial approximation. Those optimal trajectories are depicted in the left column of Figure 6.8. They are very similar for all algorithmic approaches, the joint ones as well as the two-step schemes. For guidance, we fitted a circular path with the correct radius of approximately 5.85 mm as approximate ground truth to the plot (depicted in yellow). Note that the circular path is indicated by an ellipsoidal shape as the resolution is 1 mm and 2 mm in z- and y-direction, respectively.
In a second step, we find the points of interest, i.e., the spatial positions where high concentrations are present in the first time frame. Then, we follow the motion estimates from those points on and add up motion estimates for each time step, yielding a piecewise linear trajectory depicted in the middle column of Figure 6.8. Here, several differences are already visible. Motion estimates based on images reconstructed by Kaczmarz method are too small in magnitude probably due to the severe noise artifacts in the image sequence, resulting in too narrow trajectories. The joint approaches in general perform better with visually most convincing results achieved by L1OF and Joint reconstruction algorithm using an L^2-data discrepancy term and the mass conservation motion model (L2MC).
In a third step, we smooth the results from the middle column by a fourth order polynomial again. In Table 6.5, we state the computed centers (y, z) and radii r of the best-fit circles fitted to the coordinates obtained by the piecewise linear trajectories in the middle column of Figure 6.8. The computed values confirm the visual impression, i.e., motion fields based on the Kaczmarz method severely underestimate the magnitude of motion. The reference position for the yellow circles is given by $(y_{\text{ref}}, z_{\text{ref}}) = (12.5, 13)$. Please note that the position of the reference is not exact but only a guess based on the reconstructed images. The reference circle center is closest to the values obtained for the L2MC algorithm. The Joint reconstruction algorithm using an L^2-data discrepancy term and the optical flow motion model (L2OF) center is also reasonably close. For the L^1-constrained algorithms, we observe a shift to the right. This can be explained by different means. First, the resulting image sequences are sparser, such that no averaging over different coordinates occurs which could shift the mean value. Second, several of the latter motion estimates are quite small. As the phantom moves to the right first and then to the left, this mainly affects the position of the phantom at later time steps. We obtain the best results for L1OF, slightly underestimating the magnitude of motion, and L2MC, slightly overestimating the magnitude of motion.

Table 6.5.: Centers (y, z) and radii r of the best-fit circles fitted to the coordinates obtained by following the motion estimates on each time step for the 3 Hz rotation phantom dataset. The reference value is an approximate value based on visual inspection of the image sequences combined with knowledge from the experimental setup.

	Reference	Kaczmarz	SPDHG	L1OF	L2OF	L1MC	L2MC
y	12.5	13.92	13.36	14.04	12.70	14.10	12.44
z	13	12.48	14.76	13.09	13.82	12.75	13.19
r	5.75	3.39	4.96	5.36	5.28	4.80	6.49

In Section A.2.2 in the appendix, we obtain similar results for the 7 Hz dataset in Figure A.1 and Table A.7. The more challenging image reconstruction task leads to severe

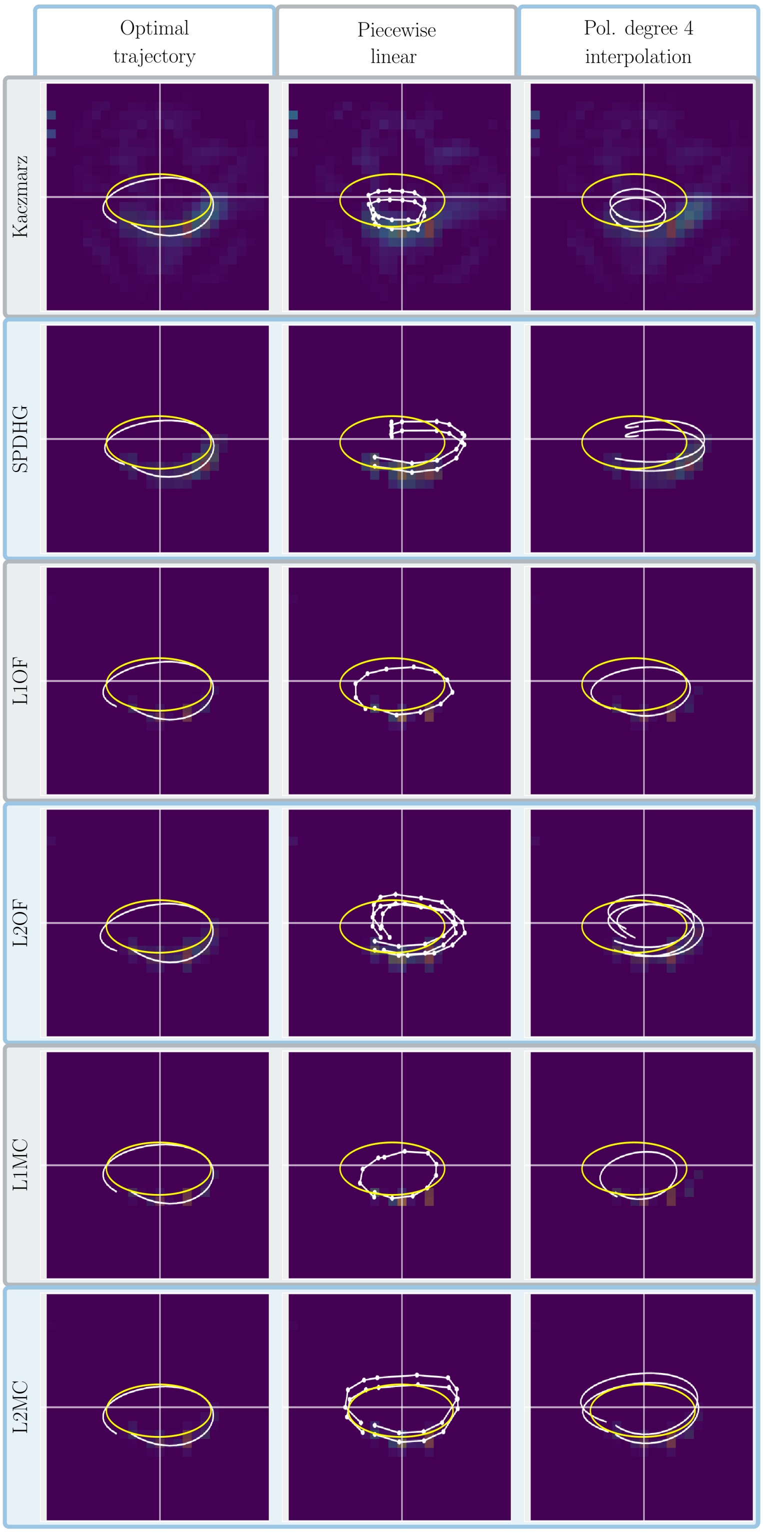


Figure 6.8.: Trajectories of dynamic particle concentration based on motion estimates by various algorithmic approaches for the 3 Hz rotation phantom dataset. The left column displays the optimal particle trajectory as observed in the reconstructed image sequence for the 3 Hz dataset by following the maximum concentration value along the frames and smoothing the result by a 4th order polynomial approximation in white. For guidance, we fit a circle with correct diameter to the curve in yellow. The middle column shows the piecewise linear trajectories, which are obtained by adding up the computed motion of every time step. The right column displays a 4th order polynomial approximation to the curve in the middle. L1OF and L2MC yield the best results, whereas the motion estimates based on the Kaczmarz-reconstructed images severely underestimates the magnitude of the displacement.

problems in the motion estimation task based on the images reconstructed by Kaczmarz method. All joint approaches yield more reasonable guesses for the trajectory, best results are obtained for the L1OF and L2MC algorithm.

The reconstruction of the rotation phantom datasets for different speeds shows that the image reconstruction task significantly benefits from the joint approach. The more appropriate priors (in comparison to the Tikhonov regularization used by Kaczmarz method) minimize noise artifacts in the reconstruction. Moreover, the reconstructed mass over the time frames is more constant due to the additional conservation prior. The motion estimates are also improved by the joint approach. The enhanced image sequences lead to significantly more accurate motion estimates.

6.3. In-vivo mouse data fulfilling the mass conservation constraint

The in-vivo mouse data were first published in the context of presenting a highly sensitive gradiometric receive coil, see [67]. Similar to the rotation phantom data, the data were measured by the Bruker Biospin scanner at the University Medical Center Hamburg-Eppendorf. In order to capture the fine structures of the cardiovascular system of the mouse, the system matrix has been calibrated with a small capillary of size 0.7 mm. This significantly higher spatial resolution is obtained by a considerably smaller delta probe compared to the previous experiments, such that we expect stronger noise on the measured signal and the measured system matrix. The system matrix captures a volume of $32.2 \times 25.2 \times 13.3\ \mathrm{mm}^3$ by $46 \times 36 \times 19$ voxels. The measurement sequence is collected by the above-mentioned gradiometric receive coil for an injected bolus of volume 10 µL. The temporal resolution is given by 21.54 ms per frame, as no time averaging was applied. The expected flow during the inflow of the bolus is about $5\,\mathrm{cm\,s^{-1}}$, see [84], which is equivalent to a movement of about 1 mm per frame. Thus, we already expect motion-related inconsistencies such that applying a motion model is beneficial. Here, we use the same data pre-processing as for the rotation phantom data with an additional correction for background signal. Therefore, we take empty measurements at the beginning of the measurement process and when calibrating the system matrix. These empty measurements are then subtracted from the phantom measurements and system matrix, respectively [136]. MRI scans provide us with background information on the structures of the inner tissue. We fit those as grayscale background images behind the reconstructed MPI image sequences.

For the motion estimation task in this section, we do not apply TV regularization as before as this yields too smooth motion estimates. The blood flow through the cardiovascular system cannot be assumed to fulfill such strong smoothness assumptions. Instead we apply L^2-Tikhonov regularization on the flow fields, which yields the optimization problem

$$\min_v \int_0^T \beta \left\| v\left(\cdot,t\right)\right\|^2_{L^2(\Omega)^n} + \gamma \left\| \frac{\partial}{\partial t} c\left(\cdot,t\right) + \nabla \cdot \Big(c\left(\cdot,t\right) \cdot v\left(\cdot,t\right)\Big) \right\|_{L^1(\Omega)} \mathrm{d}t.$$

Algorithmically, the dual update step in Algorithm 3 simplifies as the new Tikhonov term

belongs to the primal problem. The primal update step then consists of a different but easy to compute proximal operator. The regularization parameters chosen for the L1MC algorithm are $\alpha_1 = 0.05$, $\alpha_2 = 100$, $\beta = 0.1$ and $\gamma = 100$. We only present results for this algorithm for the sake of brevity.
An exemplary reconstruction of the first two frames is depicted in Figure 6.9. From left to right, the rows show different slices of the reconstructed 3D volume.

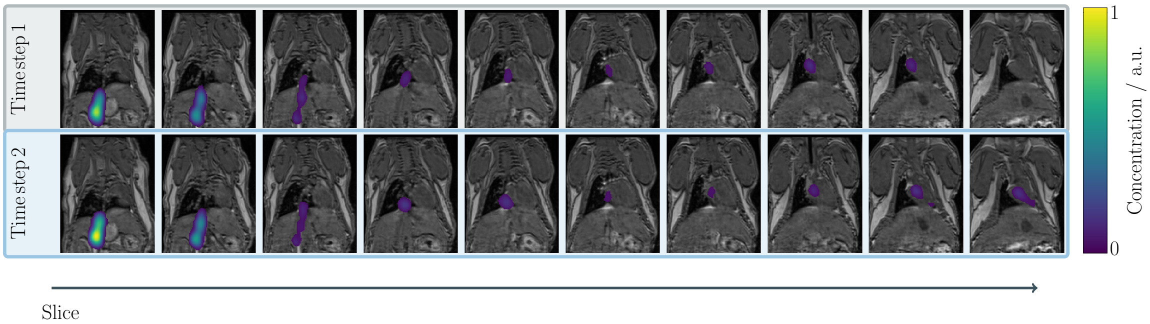


Figure 6.9.: Exemplary reconstructed image sequences from the in-vivo mousedata. Each row represents one frame in time and slices of the 3D volume are shown from left to right. The figure was first published in [27].

In Figure 6.10, we depict a single slice (index 14, corresponding to a height of 9.1 to 9.8mm) of the reconstructed volume in the left column. The motion estimates are illustrated by the quiver plots in the middle column. We observe a turbulent flow of different magnitude throughout the image section. The direction of the flow can be observed more easily by the color wheel plot in the right column, where each color represents a specific angle. Succeeding time steps are presented in the rows. The images as well as motion estimates seem reasonable, although we cannot assess the quality in comparison to a ground truth. The motion estimates depicted in Figure 6.10 do only display the motion with respect to two spatial dimensions.
In order to investigate the flow through the different slices of tissue, we consider Figure 6.11. Although this figure is difficult to interpret, we can state that the flow is reasonable to the effect that flow is visible only where concentration exists. Moreover, regions with flow directing upwards and downwards propagate over the slices, which also seems reasonable.

6.4. The subframe reconstruction setting

The high temporal resolution of MPI is one of its main advantages and outperforms many medical imaging modalities. For some applications, even a temporal resolution of 46 frames/second might not be fast enough. Thinking about blood flow imaging, the mean flow velocity in the aorta is approximately $12\,\mathrm{cm\,s^{-1}}$, which translates to 2.6 mm/frame in a standard 3D setup. In the arteries, even velocities up to $45\,\mathrm{cm\,s^{-1}}$ are possible, which is equivalent to 9.7 mm/frame, see [58]. Thus, large displacements are possible within the repetition time of an MPI scanner.
If we then additionally consider the fact that motion estimation results in non-unique solutions and that we favor solutions, which represent the shortest way by our standard

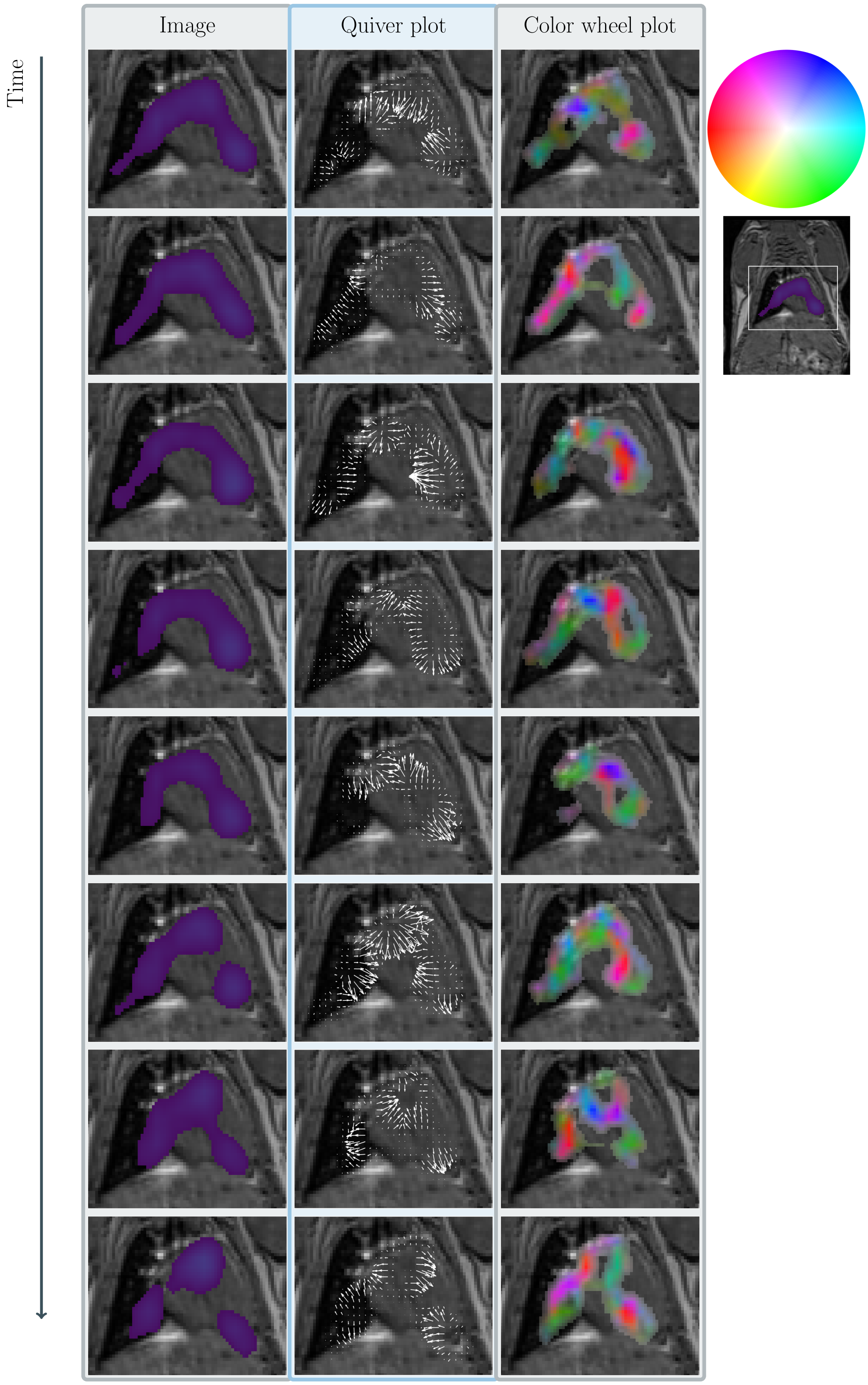


Figure 6.10.: Reconstructed images and flow for the in-vivo mousedata. Reconstructed images for eight time frames are depicted in the left column, a quiver plot of the corresponding motion estimates in the middle column and a color wheel plot indicating the direction of motion in the right column. The figure was first published in [27].

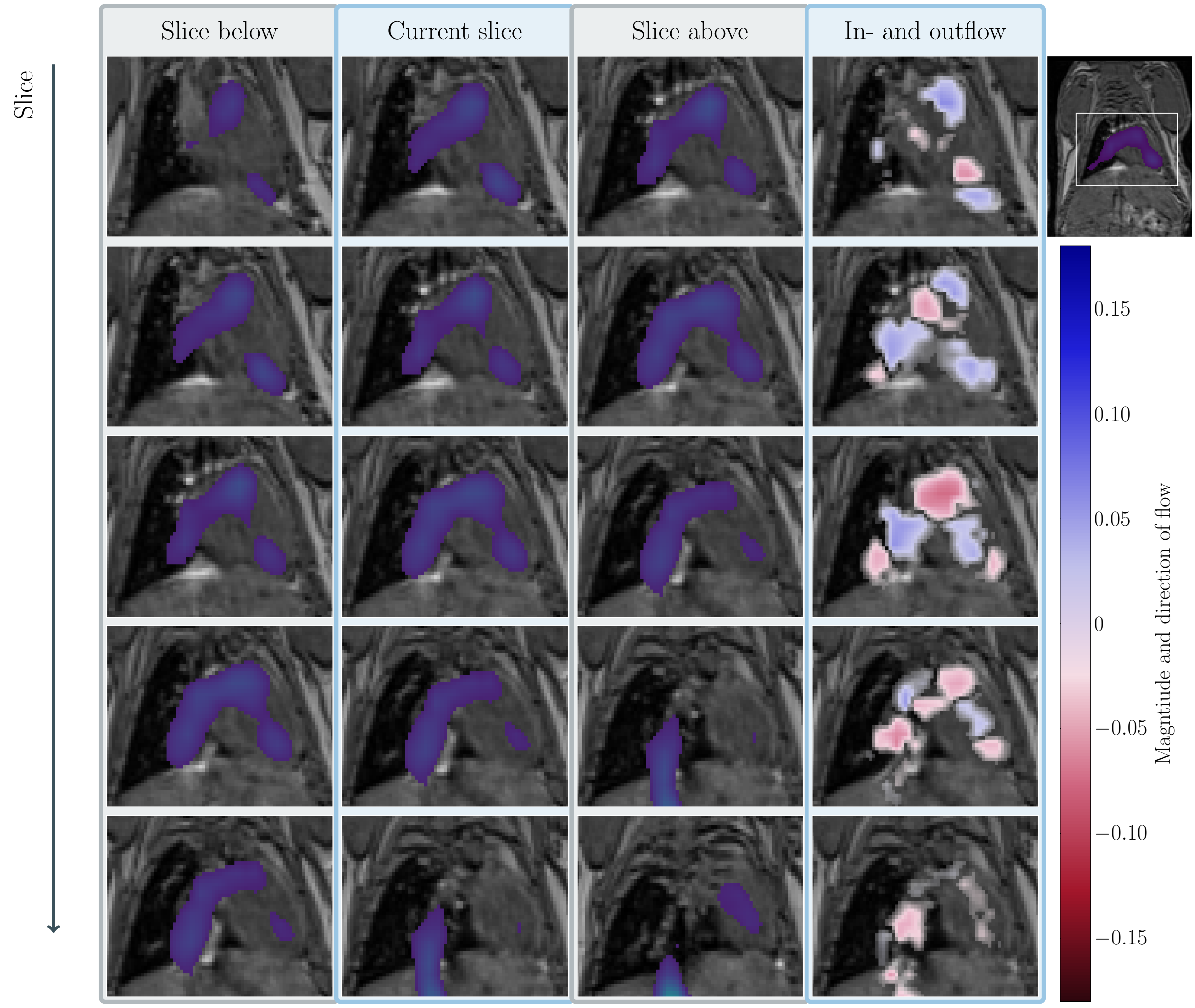


Figure 6.11.: Reconstructed images and flow in z-direction for the in-vivo mousedata for one time step. The second left column displays the current slice of the 3D volume, the first column depicts the slice below and the third column shows the slice above. In the right column, the in- and outflow into the current slice is depicted. Blue areas indicate flow upwards (to slice above), red areas indicate a downward flow (to slice below). The figure was first published in [27].

algorithmic schemes, we might not be able to investigate the exact flow and the dynamics if we restrict ourselves to the repetition time for the temporal resolution. However, it might be of high diagnostic relevance to observe the exact way of the blood flow and see whether the flow goes through or around an obstacle. A natural idea to improve the temporal resolution of the reconstructed image sequences is to derive forward operators and measurements for shorter time intervals. This is possible due to the measurement process of MPI, cf. Figure 3.3. Remember that we denoted the repetition time of an MPI scanner by T_R. In the following, we define the repetition time of a subframe by T_{sf}. The data u used for reconstruction of one subframe is obtained during an interval of length T_{sf}, i.e.,

$$u\,(t) = \int_\Omega c(x,t)s(x,t)\mathrm{d}x \qquad \text{for } \ t \in (\, 0, T_{\mathrm{sf}}\,]\,. \tag{6.1}$$

This data sampling during subintervals is illustrated by Figure 6.12. The FOV is sampled densely during a full frame, but only partly and not as dense during subframes. The more subframes we consider, the better is the temporal resolution but the worse is the spatial resolution. If we use other excitation functions such that a non-Lissajous trajectory occurs, this problem might be intensified.

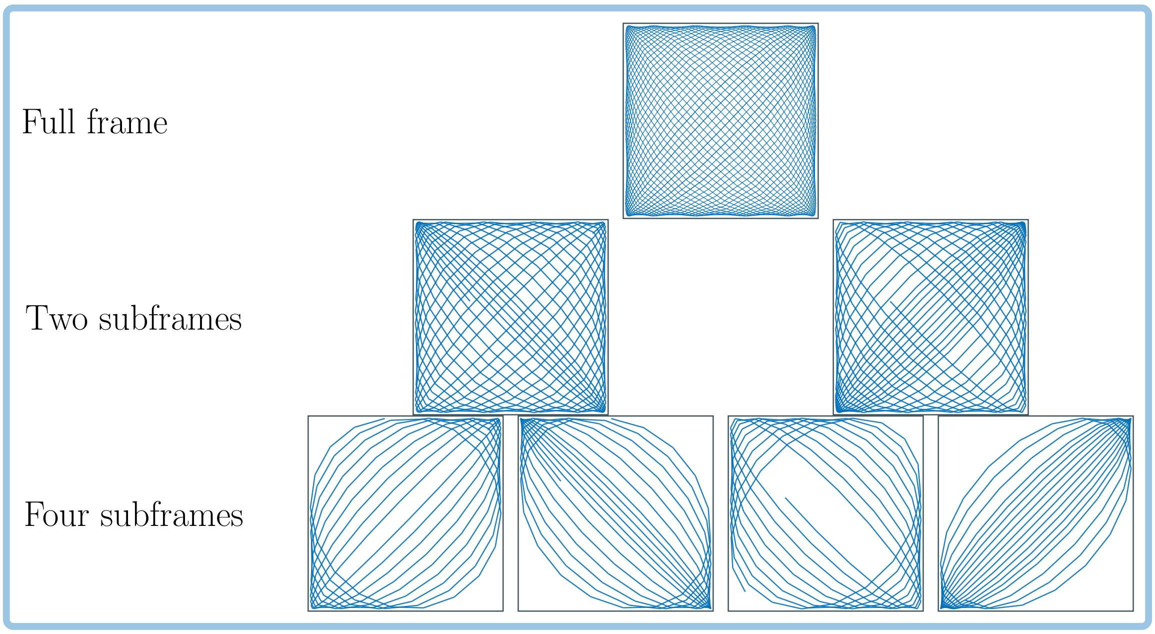


Figure 6.12.: The field-free point trajectories cover only parts of the field-of-view (FOV) for subframe reconstruction. We depict a Lissajous trajectory that samples the 2D FOV fine-mesh during a full frame (upper row). In the middle row, we consider the subframe reconstruction setting with $T_{\mathrm{sf}} = T_R/2$, i.e., the repetition time of a subframe corresponds to half of the repetition time of a full frame. In the bottom row, we illustrate the trajectories for the subframe setting with $T_{\mathrm{sf}} = T_R/4$. Large areas are then not scanned sufficiently during a subframe.

Remark. All results regarding the existence of minimizers to the joint optimization problem still hold in the subframe reconstruction setting as long as the choice of excitation function still ensures a point-wise non-vanishing forward operator.

Main consequences of subframe considerations are the temporally higher resolution, but also the more limited data available for each image reconstruction problem. Moreover, the trajectories of the FFP are not closed anymore. As mentioned before, we expect a spatially lower resolution, as parts of the FOV may not be covered at all by the FFP trajectory. Another side-effect relates to the reconstruction space. Previously, we performed image reconstruction in the Fourier domain. Now, we consider the data in the Fourier domain to perform a frequency selection, but then transform it back to the time domain. Doing this after frequency selection naturally incorporates an additional error, but we need the frequency selection in order to achieve a reasonable SNR. However, we need to reconstruct in the time domain in order to perform the temporally correct grouping of the data. This describes our data pre-processing already, additionally we normalize the system matrices and corresponding measurements.

In the following, we analyze subframe reconstruction on the simulated dataset, cf. Section 6.1. We restrict ourselves to the L1OF algorithm for the joint approach for brevity. This algorithm is one of the two best-performing joint schemes on the simulated data as we showed in Section 6.1. We compare its performance to the static reconstruction schemes based on the Kaczmarz method and SPDHG. Again, we consider the three different noise levels noise-free, medium noise and high noise.
First, we set $T_{\text{sf}} = T_R/2$, i.e., one original temporal frame consists of two subframes such that the temporal resolution is twice as high as before. In a noise-free scenario, the static approaches by Kaczmarz method and SPDHG perform well. In this setting, the reconstruction of the phantom is reasonable in size and position. However, already in the medium noise scenario, the Kaczmarz method reconstructs image sequences suffering from severe noise artifacts. The position of the phantom is not clearly distinguishable from the background. SPDHG now also produces image sequences with artifacts but still yields good results. The image sequence reconstructed by the joint L1OF algorithm certainly has the highest quality, as the phantom is in the correct position and the background is empty. In the high noise scenario, the joint L1OF algorithm outperforms the two static approaches. The image sequence reconstructed by the Kaczmarz method is not recognizable anymore, the images by SPDHG suffer from severe artifacts which cannot be distinguished from the phantom. The joint L1OF yields image sequences with the phantom in the correct position and with only slight artifacts regarding its shape. For illustration, all reconstructed image sequences for the first four subframes are depicted in Figure 6.13.
Now, we are interested in achieving an even higher temporal resolution. Again, we halve the interval length, such that $T_{\text{sf}} = T_R/4$. This yields a maximum displacement of one voxel for each time step. However, on this simulated data set, we have only 800 observations left to reconstruct the concentration values of 8.000 voxels. We reconstruct an image sequence by L1OF. In a noise-free setting, no artifacts in the reconstruction are visible. In the medium noise scenario, we observe slight artifacts clearly distinguishable from the phantom. However, in the high noise scenario stronger artifacts exist. Still, the phantom is clearly distinguishable from background and noise. Reconstructed image sequences for all noise levels are illustrated in Figure 6.14. We increase the temporal resolution further to $T_{\text{sf}} = T_R/8$. In this case, we fulfill the assumption of a quasi-static tracer distribution such that an even finer temporal resolution is not useful. Figure 6.15 depicts reconstructed image sequences showing the good performance of the joint algorithm even in this

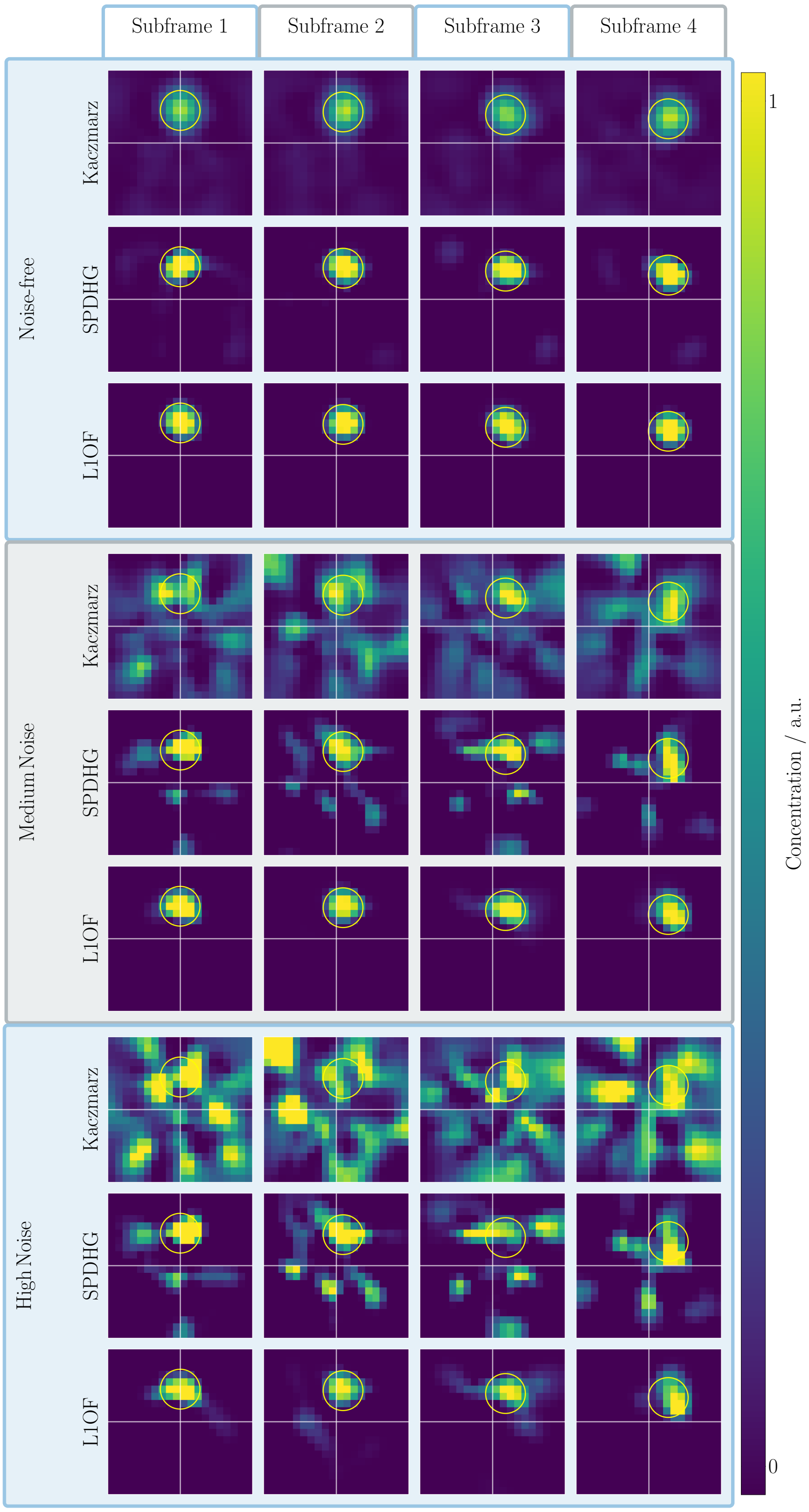


Figure 6.13.: Subframe reconstructions by different algorithmic approaches for simulated data. One original frame corresponds to two subframes, i.e., $T_{\text{sf}} = T_R/2$. On noise-free data (upper block) the static methods as well as the joint approach reconstruct the phantom in the correct size and position. For the medium noise level (middle block), the phantom is still recognizable when reconstructed by static SPDHG. In the image sequence reconstructed by Kaczmarz method the background noise cannot be distinguished from the phantom anymore. For the high noise level (bottom block) only the joint L1OF yields convincing results. The images reconstructed by static approaches suffer from noise and motion artifacts. The true position of the phantom for each subframe is indicated by the yellow circle. All image sequences are depicted as projections onto the x-y-plane.

case. Neither the static Kaczmarz method nor the static SPDHG approach are able to produce meaningful results in this setting.

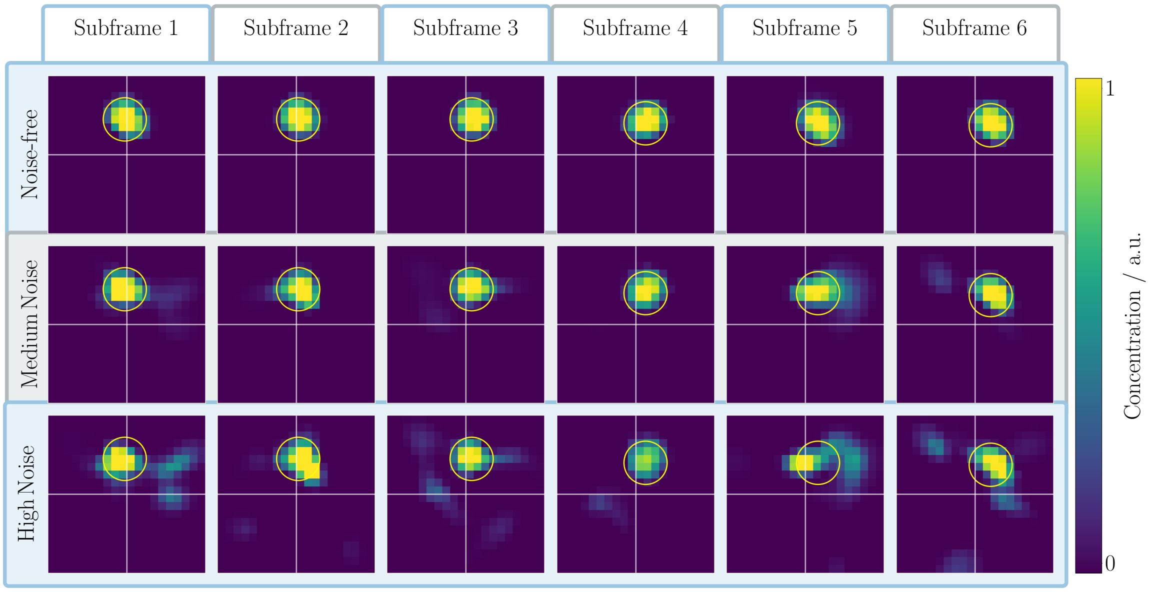


Figure 6.14.: Subframe reconstructions by the joint L1OF algorithm for quarter frames, i.e., $T_{\text{sf}} = T_R/4$ for three different noise scenarios for simulated data. In the noise-free setting (top row), the phantom is well resolved. Slight artifacts exist in the medium noise setting (middle row) and become stronger in the high noise scenario (bottom row). The true position of the phantom for each subframe is indicated by the yellow circle. All image sequences are depicted as projections onto the x-y-plane.

In a next step, we consider the effect of subframe reconstruction to the motion estimation task in the joint approach. The higher temporal resolution leads to smaller displacements between the different image frames. This simplifies the numerical motion estimation, as large displacements are typically more difficult to assess correctly. Moreover, larger displacements between two original frames are estimated by a linear displacement. By applying subframe reconstruction, this displacement splits into several smaller linear translations. As a whole, they describe a more complex shift. On the downside, small errors on each motion estimate add up to a larger one, as a greater number of motion estimates is computed overall.

Figure 6.16 depicts the computed displacements over the first three frames. In this example, we use the L2MC algorithm. The ground truth positions for the first frame are colored in blue, for the second one in red and for the third one in gray. We indicate the path via a solid line although the phantom does not move continuously. The mean position of the full frame is indicated by a large circular mark and the mean positions of the four subframes are depicted by smaller circular marks. We see as a direct consequence that connecting the full frame mean positions directly does not match the result when connecting the mean positions during quarter frames. Thus, information about the particles' path gets lost if we perform full frame reconstruction.

The computed motion estimates for full frames (a) and quarter frames (b) are indicated by arrows in Figure 6.16. Small inaccuracies in each subframe in (b) yield an estimated trajectory which is slightly shifted. We observe small as well as large errors for the subframe motion estimates, paying tribute to the input image sequences which suffer from

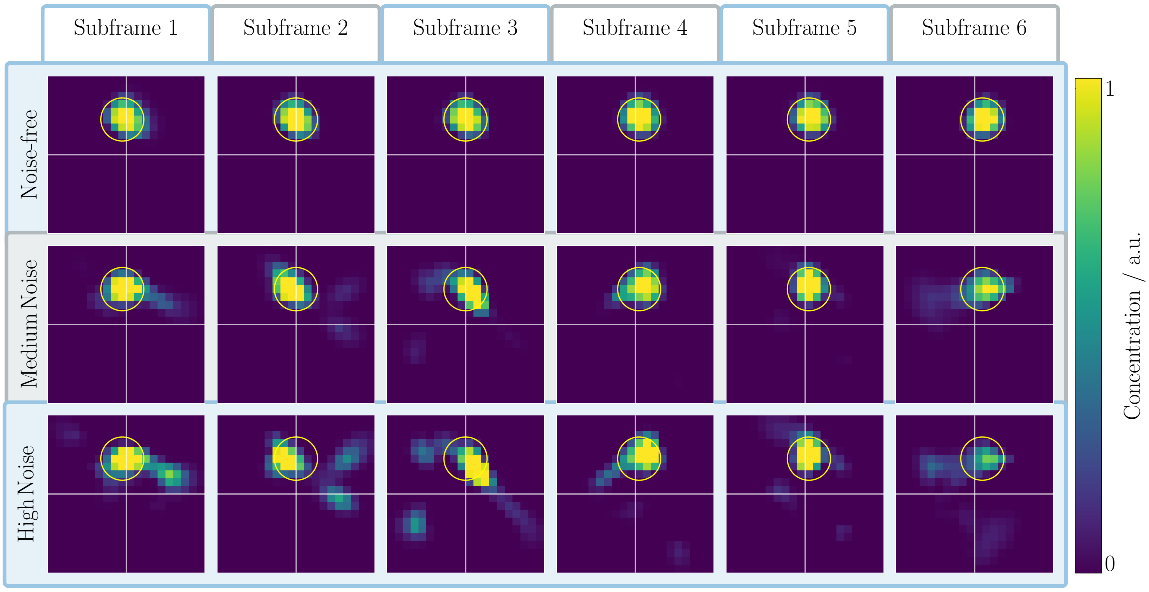
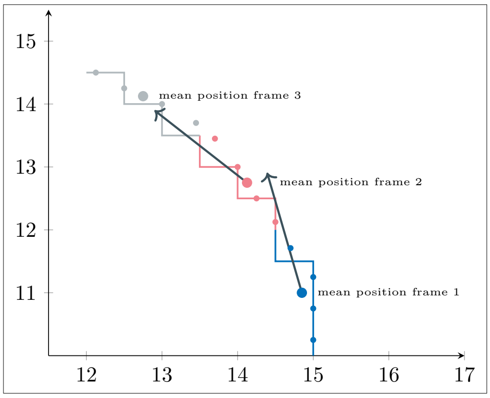
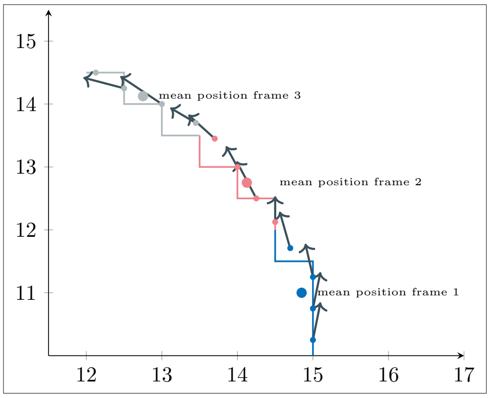


Figure 6.15.: Subframe reconstructions by the joint L1OF algorithm for eighth frames, i.e., $T_{\mathrm{sf}} = T_R/8$ for three different noise scenarios for simulated data. In the noise-free setting (top row), the phantom is well resolved. In the medium noise scenario (middle row), we can still clearly distinguish the phantom from the background, although its shape is slightly distorted. In the high noise scenario (bottom row), strong artifacts exist in some frames making it difficult to spot the exact position of the phantom. The true position of the phantom for each subframe is indicated by the yellow circle. All image sequences are depicted as projections onto the x-y-plane.



(a) Full frame displacements.



(b) Quarter frame displacements.

Figure 6.16.: Comparison of full frame and subframe motion estimates using the L2MC algorithm. The true position of the phantom during the frames is indicated by the colored lines (blue indicates the first frame, red the second one and gray the third one, respectively). The mean position during the frames is indicated by the larger circular marks, the smaller ones describe the mean positions during subframes. In (a), the motion estimates in the full frame reconstruction setting are indicated by the arrows. In (b), we illustrate the motion estimates in the subframe setting. The particles' path can be tracked with a lot more detail for the subframe case. However, the higher number of estimates comes at the cost of several small errors that add up to a larger one.

artifacts. Still, the direction of motion is more accurately estimated when considering the subframe approach instead of the full frame displacements. We can well observe how the direction changes gradually during the sequence. By using larger subframes or full frames, information is lost due to the averaging effect occurring when considering data from longer time spans.
The parameter γ which weights the motion model penalty term is especially important in the subframe setting. It weights the influence of the motion model and thus links the motion estimation task and the image reconstruction task. The image reconstruction task is quite robust to the choice of γ, for the motion estimation task we observe the behavior illustrated in Figure 6.17. For smaller parameters γ, the computed direction of the displacement differs more strongly from the true direction. This behavior meets our expectations, as by using a smaller γ we allow for larger discrepancies to the motion model. However, we observe a maximum value that should not be exceeded in order to gain a reasonable output. For larger γ, in this example larger than $\gamma = 10^{-2}$, the displacement fields point in random directions and do not match the true displacement. In this case, the regularization within the motion estimation subproblem is not strong enough compared to the data term, i.e., the motion model.

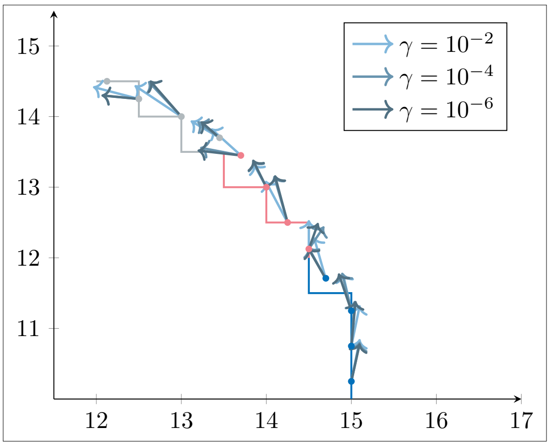


Figure 6.17.: The influence of the parameter γ on the motion estimation task in a subframe setting. The direction of motion is computed more accurately for larger but sufficiently small γ. If γ is chosen larger than 10^{-2} the algorithm fails to compute a reasonable solution.

We conclude the consideration of the subframe setting by showing the applicability to measured data. Therefore, we consider the 3 Hz rotation phantom from Section 6.2. We use the same data preprocessing, additionally splitting data and matrices into subframe intervals. The regularization parameters are set similarly to the full frame setting, cf. Table 6.4. These parameters are not tuned but simply adopted from the full frame setting. Figure 6.18 illustrates the reconstructed image sequence by projections onto the different planes. The subframe results describe the position of the phantom with high precision and thereby achieve a higher temporal resolution compared to the fullframe image sequence. However, the size of the phantom is distorted in x-direction. This might be due to the non-optimal regularization parameters on the one hand and due to the scanning trajectories which are not dense enough while applying subframe reconstruction on the other hand. The effect amplifies if we consider quarter frames instead of half frames.

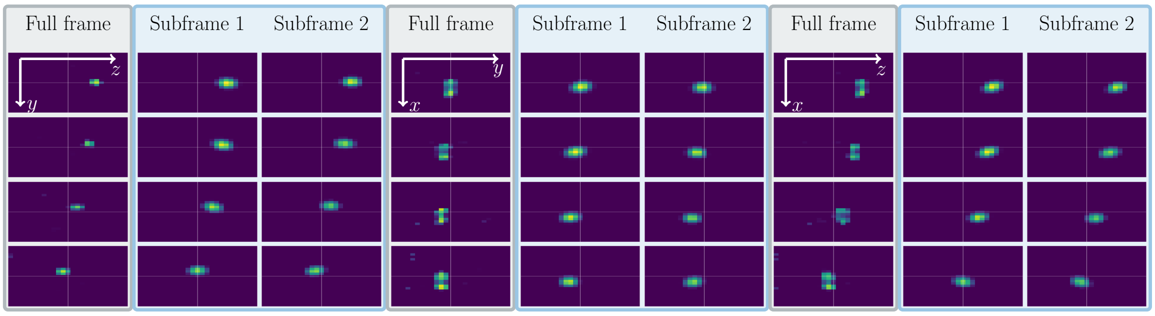


Figure 6.18.: Comparison of subframe ($T_{\mathrm{sf}} = T_R/2$) and full frame results for the rotation phantom. While the position and the size of the phantom in z- and y-direction are well resolved, the size in x-direction is too small for the subframe results. However, the spatial position is captured with higher resolution compared to the full frame setting.

6.5. Summary and discussion

In this section, we tested the proposed joint reconstruction approach for image sequences and motion estimates numerically. We first showed the superior results regarding the image quality compared to standard static reconstruction approaches taking into account different measures for the image quality on synthetic data in Section 6.1. Thereby, we considered not only ideal conditions but also a noise-corrupted scenario. We analyzed the main drawback of the proposed method, i.e., the high computational costs of certain aspects of the joint algorithm, in particular of the image reconstruction task. The computation of the inverse motion field for the OF constrained algorithms is costly but has the potential for a speed-up of the algorithm by a more efficient calculation of the inverse motion field and by cheaper computations of differential operators.

After these first proof-of-concept results on simulated data we considered measured data, namely the rotation phantom in Section 6.2. Based on visual inspection, we again observed high quality results by the proposed joint approach. By evaluating the trajectories obtained from the motion estimates, we underlined the good performance of the approach with respect to motion estimation task.

The third data set considered was in-vivo data imaging the cardiovascular system of a mouse in Section 6.3. This data fulfill only the MC constraint but not the OF constraint. Imaging blood flow in the cardiovascular system, we used a regularization term assuming less smoothness of the flow compared to a TV regularization term. More precisely, we implemented an L^2-Tikhonov regularization on the motion fields. The results are reasonable based on visual inspection, showing the applicability of the proposed approach also to more complex scenarios and especially to real-world problems.

In Section 6.4, we considered a subframe reconstruction setting. In this setting, we enhanced the temporal resolution of the computed image sequences and motion fields. To achieve this, we artificially reduced the length of a measurement frame to a subframe. This leads to less data available per subframe and measurements with a lower spatial resolution, posing an overall more difficult image reconstruction task. However, we showed that the proposed joint approach benefits from linking the images at different time steps and yields superior results compared to standard static techniques. We first considered simulated data, where we observed that static methods can solve the task for half frames in a noise-free setting. When introducing noise as well as when further increasing the temporal resolution to quarter or eighth frames, only the proposed joint approach produces convincing results. We closed the section by applying the subframe approach to the 3 Hz rotation phantom data from Section 6.2, briefly showing the applicability also on measured data.

We derive different possible directions for future research from the numerical performance of the algorithms. In order to test the accuracy of the motion estimation task from measured data, we need to carefully design a dynamic in-vitro phantom. For this phantom, we need simple flows in order to understand flow directions and magnitude at all spatial positions from limited measurements, e.g., measurements of the inflow speed when inserting tracer material.

Various extensions like, e.g., allowing for consideration of in- and outflow into the FOV, multi-patch setups and inclusion of prior knowledge from background images has to be investigated from the theoretical as well as numerical side.

In order to make the runtime competitive and ideally enable online reconstruction, we have to find a more efficient way to invert the motion fields. Moreover, the matrix-vector products that are very costly in MATLAB might be less costly when implemented in other languages. Further savings in terms of runtime might be possible by porting the differential operators, e.g., to C ++.

7. Conclusion

7.1. Summary

In this thesis, we considered joint image reconstruction and motion estimation for three-dimensional dynamic MPI from a theoretical as well as a numerical perspective.
In Chapter 3, we analyzed the static MPI forward operator in detail. We showed compactness of the static three-dimensional forward operator on compact time intervals, yielding the ill-posedness of the image reconstruction task. With the help of an auxiliary lemma, we verified a regularity property of the MPI forward operator. More precisely, we considered 3D MPI with continuously differentiable trajectories of the FFP in a neighborhood of a nonstationary time point of the FFP trajectory. We showed that the forward operator is then non-vanishing in this neighborhood. We briefly studied the full dynamic forward model instead of the static one and observed different temporal scales related to the dynamic imaging task. However, noting that the full dynamic model is computationally too complex to be incorporated as a forward operator into our scheme, we restricted ourselves to the assumption of quasi-static tracer distributions. To conclude the chapter, we explained different image reconstruction algorithms, namely the current state-of-the-art Kaczmarz method and SPDHG, a primal-dual approach allowing for different regularization terms. Comparing those approaches on synthetic data, we observed that Kaczmarz method yields blurry and noise-corrupted reconstructions, whereas SPDHG yields cleaner image sequences. This fact is particularly important, as having clear and clean input images is crucial for successfully applying a motion estimation algorithm.
In Chapter 4, we dealt with the motion estimation task in general as well as an application to MPI-like images. We gave an introduction to the motion estimation problem and briefly considered classical methods, e.g., Horn-Schunck and Lucas-Kanade. We derived two motion models, namely OF and MC, which we incorporated into our algorithmic approach as data fidelity terms. We used the primal-dual PDHG algorithm in order to solve the corresponding optimization problems and found that the different motion models yield slightly different algorithmic schemes. A Taylor expansion and the discretization of the differential operators limit the maximum magnitude of motion which can be detected. We proposed a multiscale and warping scheme for the OF motion model and a multiscale scheme for the MC motion model in order to be able to recover large displacements. The resulting algorithms were tested on phantoms that resemble MPI images in the sense that they are very sparse and have an empty background. This poses a very specific motion estimation task, such that only such images are considered in this work. We showed that the multiscale schemes significantly improved the motion estimation when applied to displacements exceeding one voxel per frame. Moreover, it depends on the specific setting whether the OF or the MC based algorithm yields better results. Extending our algorithms, we also applied gaussian smoothing on the input images in order to obtain a

higher robustness to noise. However, this did not improve the overall outcome significantly and was only applicable to the OF based algorithm. We then proposed a structural prior based on background images and their edges. This prior aims at promoting flows which are likely to occur from a physical perspective and suppress flows which are unlikely to occur. More precisely, flows aligning with edges in background images are promoted whereas flows perpendicular to such edges are suppressed. This prior was only applicable to the MC based algorithm and increased its angular accuracy under noise significantly. Concerning the run time and cost, we saw that computation of spatial derivatives is the main cost factor within the PDHG algorithms and iterations on finer scales are very costly compared to lower scales. We considered a stopping criterion based on the objective value to limit the number of iterations on the finest scales, resulting in a small save of run time.
In Chapter 5, we finally considered the joint image reconstruction and motion estimation task. Starting from the theoretical perspective, we showed existence of a minimizer to the joint problem incorporating the motion model as a hard constraint almost everywhere in Theorem 5.1. We followed the proof of [35, 50], lifting the dimensionality from two to three dimensions and including the time-dependent forward operator. Dimension dependent embeddings were used at crucial points and had to be adapted which was the main challenge of the proof. In the following Lemma 5.1, we proved convergence of solutions of the unconstrained minimization problems incorporating the motion model as an additional penalty term to a solution of the constrained minimization problem in terms of Γ-convergence. Moreover, we briefly commented on the regularization property which the proposed joint approach comprises. We considered an alternating minimization approach to solve the joint problem and briefly discussed convergence of such an algorithm.
The numerical evaluation of the proposed joint image reconstruction and motion estimation algorithm is contained in Chapter 6. We started by experiments on simulated data, showing the superior performance of the joint approach compared to a two-step scheme in terms of SSIM and PSNR for the image reconstruction task. Moreover, we observed that only few joint iterations are needed until only negligible changes occur for the reconstructed images as well as the motion estimates. Afterwards, we considered real measured data instead of synthetic data and showed the applicability of our approach also on the OF-like rotation phantom data set and the MC-like in-vivo mouse data set. We obtained improved motion estimates by the joint approach compared to a two-step scheme as well as higher perceived image quality. We closed the numerical evaluation by an investigation of the subframe setting. In this setting, we enhance the temporal resolution of the reconstructed image sequences by temporally different data grouping. We introduced subframes which are defined as partial frames and reconstructed images on this temporal scale. Therefore, we had to reconstruct in the time domain instead of the Fourier domain which was previously used. Starting again on simulated data, we showed superior performance of the joint approaches compared to static reconstruction methods. Especially in a noise-corrupted setting, the proposed joint approach benefited from the temporally coupled formulation and was able to produce high-quality results, whereas static methods did not manage to recover the phantom.

7.2. Discussion and outlook

Concerning the motion estimation task, we pointed out that the sparse, empty background image sequences pose a very specific challenge. We found that application of our primal-dual motion estimation algorithm yields fast degrading quality of the displacement fields if applied to noisy data. It is necessary to derive a more robust motion estimation algorithm tackling this problem. The proposed structural prior is a first step in that direction but needs to be extended to the OF based motion estimation algorithm. Another option might be incorporating a temporally smoothing regularization term. Instead of applying the motion estimation task to the reconstructed image sequences, one could also try to retrieve motion information directly from the raw measured MPI signal, i.e., from information in the frequency domain. Moreover, the currently used algorithm could potentially be improved by a tailored stopping criterion on the different scales, optimized step size strategies and more refinement steps per scale. From the implementation side, there is room for improvement of the efficiency, potentially by porting costly parts of the script to different languages.

Considering now the joint approach from a theoretical perspective, we see that within the formulation of the unconstrained joint problem, the motion model defines an additional part of the, now nonlinear, forward operator. Under strong assumptions, it was shown straightforward that the joint problem formulation defines a regularization method. Under more general assumptions, however, it is a challenging task and left for future work. We applied the proposed joint approach to a specific MPI device, working in three dimensions, having linearly independent receive coil units and a field-free point geometry fulfilling a continuously differentiable trajectory. These assumptions are not restrictive in general, as MPI is intrinsically three dimensional and the receive coils are typically even orthogonal. Moreover, a Lissajous trajectory fulfills the assumptions for sinusoidal excitation functions. An extension of the setting of Theorem 5.1 would be necessary for cosine excitation functions and for different scanning geometries like a field-free line scanner or single sided scanners.

Note that we used the static forward model throughout our considerations which assumes at least quasi-static tracer distributions. In order to limit the modeling error, we need to extend the setting to the full dynamic forward model. To date, it is not possible to measure the full dynamic forward model and thus a (at least partly) model-based approach for the system matrix has to be applied. This again poses a research question on its own which is why we do not consider it in this work. However, it would be even more interesting to analyze the behavior of the full dynamic model and check whether it fulfills the same regularity assumption.

In this work, we only briefly touched the topic of subframe reconstruction. Discovering the challenges of subframe reconstruction in more detail is an interesting research question which is left for future work. Especially the impact of the not-as-dense sampling during the measurement process has to be studied. Our numerical results on the rotation phantom data set showed difficulties in retrieving the original size and shape of the phantom, this might be due to a spatially not sufficient sampling process but needs to be understood in detail. It might be useful to consider additional temporal smoothness when applying a

subframe approach in order to link consecutive time steps.
Moreover, we briefly introduced the multi-patch setting for MPI measurements. As the FOV in MPI is very limited, a realistic scenario in practice will always involve multiple patches. This does not pose huge challenges for the image reconstruction task alone. However, when applying a joint image reconstruction and motion estimation task with a motion model based on conservation laws, we have to carefully consider the setting. Thus, in- and outflow into certain patches have to be modeled and the numerical implementation of the conservation assumption has to be handled explicitly.
From a practical point of view, the applicability of the proposed approach to MPI image sequences was shown on measured data. The resulting image sequences were visually inspected and showed good results. However, accessing the quality of the resulting displacements field is not possible without having an approximate ground truth. Therefore, we need a carefully designed experimental setup allowing for an evaluation of the ground truth flow. We leave this exciting task for future work.

Bibliography

[1] R. A. Adams and J. J. Fournier. *Sobolev spaces.* Elsevier, 2003.

[2] M. Ahlborg, C. Kaethner, T. Knopp, P. Szwargulski, and T. M. Buzug. Using data redundancy gained by patch overlaps to reduce truncation artifacts in magnetic particle imaging. *Phys. Med. Biol.*, 61(12):4583, 2016.

[3] H. W. Alt. *Linear functional analysis: An Application-oriented Introduction.* Springer London, 2016.

[4] L. M. Álvarez León, J. Esclarín Monreal, M. Lefébure, and J. Sánchez. A PDE model for computing the optical flow. *Proceedings of CEDYA XVI.*, 1999.

[5] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems.* Oxford university press, 2000.

[6] S. Arridge, P. Fernsel, and A. Hauptmann. Joint reconstruction and low-rank decomposition for dynamic inverse problems. *Inverse Probl. Imaging*, 16(3):483–523, 2022.

[7] B. Askin, A. Güngör, D. Alptekin Soydan, E. U. Saritas, C. B. Top, and T. Cukur. PP-MPI: A deep plug-and-play prior for magnetic particle imaging reconstruction. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 105–114. Springer, 2022.

[8] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.*, 35(2):438–457, 2010.

[9] G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM J. Appl. Math.*, 60(1):156–182, 1999.

[10] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*, volume 147. Springer, 2006.

[11] J.-P. Aubin. Un théoreme de compacité. *CR Acad. Sci. Paris*, 256(24):5042–5044, 1963.

[12] L. Bar, B. Berkels, M. Rumpf, and G. Sapiro. A variational framework for simultaneous motion estimation and restoration of motion-blurred video. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, 2007.

[13] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *Int. J. Comput. Vis.*, 12(1):43–77, 1994.

[14] J. L. Barron and N. A. Thacker. Tutorial: Computing 2D and 3D optical flow. *Imaging Science and Biomedical Engineering Division, Medical School, University of Manchester*, 1, 2005.

[15] C. Bathke, T. Kluth, C. Brandt, and P. Maaß. Improved image reconstruction in magnetic particle imaging using structural a priori information. *Int. J. Magn. Part. Imaging*, 3(1), 2017.

[16] A. Beck. On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes. *SIAM J. Optim.*, 25(1):185–209, 2015.

[17] A. Beck. *First-order methods in optimization.* SIAM, Philadelphia, PA, 2017.

[18] F. Becker, S. Petra, and C. Schnörr. Optical flow. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 1945–2004. Springer, New York, 2015.

[19] D. Bertsekas. *Convex optimization theory*, volume 1. Athena Scientific, 2009.

[20] J. Bigün and G. H. Granlund. Optical flow based on the inertia matrix of the frequency domain. *Proc. SSAB Symposium on Picture Processing*, 1988.

[21] J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(08):775–790, 1991.

[22] S. E. Blanke, B. N. Hahn, and A. Wald. Inverse problems with inexact forward operator: iterative regularization and application in dynamic imaging. *Inverse Probl.*, 36(12):124001, 2020.

[23] C. Blondel, R. Vaillant, G. Malandain, and N. Ayache. 3D tomographic reconstruction of coronary arteries using a precomputed 4D motion field. *Phys. Med. Biol.*, 49:2197–2208, 2004.

[24] J. Bohnert and O. Dössel. Calculation and evaluation of current densities and thermal heating in the body during MPI. In *Magnetic Nanoparticles: Particle Science, Imaging Technology, and Clinical Applications*, pages 162–168. World Scientific, 2010.

[25] J. Borwein and A. Lewis. *Convex analysis and nonlinear optimization.* Springer, New York, 2006.

[26] A. Braides et al. *Gamma-convergence for Beginners*, volume 22. Clarendon Press, 2002.

[27] C. Brandt, T. Kluth, T. Knopp, and L. Westen. Dynamic image reconstruction with motion priors in application to three dimensional magnetic particle imaging. *SIAM J. Imaging Sci.*, 17(3):1539–1586, 2024.

[28] C. Brandt and C. Schmidt. Modeling magnetic particle imaging for dynamic tracer distributions. *Sens. Imaging*, 22(1):1–24, 2021.

[29] C. Brandt and C. Schmidt. Motion compensation for non-periodic dynamic tracer distributions in multi-patch magnetic particle imaging. *Phys. Med. Biol.*, 67(8):085005, 2022.

[30] K. Bredies and D. Lorenz. *Mathematical image processing.* Springer, 2018.

[31] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 1. Springer, 2010.

[32] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *European conference on computer vision*, pages 25–36. Springer, Berlin, Heidelberg, 2004.

[33] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *Int. J. Comput. Vis.*, 61(3):211–231, 2005.

[34] M. Burger, H. Dirks, L. Frerking, A. Hauptmann, T. Helin, and S. Siltanen. A variational reconstruction method for undersampled dynamic x-ray tomography based on physical motion models. *Inverse Probl.*, 33:124008, 2017.

[35] M. Burger, H. Dirks, and C.-B. Schonlieb. A variational model for joint motion estimation and image reconstruction. *SIAM J. Imaging Sci.*, 11(1):94–128, 2018.

[36] M. Burger, A. C. Mennucci, S. Osher, and M. Rumpf. A guide to the TV zoo. *Level Set and PDE Based Reconstruction Methods in Imaging: Cetraro, Italy 2008*, pages 1–70, 2013.

[37] M. Burger, T. Schuster, and A. Wald. Ill-posedness of time-dependent inverse problems in Lebesgue-Bochner spaces. *Inverse Problems*, 40(8):085008, 2024.

[38] T. M. Buzug, G. Bringout, M. Erbe, K. Gräfe, M. Graeser, M. Grüttner, A. Halkola, T. F. Sattel, W. Tenner, H. Wojtczyk, J. Haegele, F. M. Vogt, J. Barkhausen, and K. Lüdtke-Buzug. Magnetic particle imaging: Introduction to imaging and hardware realization. *Zeitschrift für Medizinische Physik*, 22(4):323–334, 2012.

[39] B. Cengiz. On the duals of Lebesgue-Bochner L^p spaces. *Proc. Am. Math. Soc.*, 114(4):923–926, 1992.

[40] A. Chambolle, C. Delplancke, M. J. Ehrhardt, C.-B. Schönlieb, and J. Tang. Stochastic primal dual hybrid gradient algorithm with adaptive step-sizes. *J. Math. Imaging Vis.*, 66:294–313, 2024.

[41] A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schönlieb. Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications. *SIAM J. Optim.*, 28(4):2783–2808, 2018.

[42] C. Chen. Spatiotemporal imaging with diffeomorphic optimal transportation. *Inverse Probl.*, 37(11):115004, 2021.

[43] C. Chen, B. Gris, and O. Öktem. A new variational model for joint image reconstruction and motion estimation in spatiotemporal imaging. *SIAM J. Imaging Sci.*, 12:1686–1719, 2019.

[44] C. Clason and T. Valkonen. Introduction to nonsmooth analysis and optimization. *arXiv preprint arXiv:2001.00216*, 2020.

[45] C. Crawford, K. King, C. Ritchie, and J. Godwin. Respiratory compensation in projection imaging using a magnification and displacement model. *IEEE Trans. Med. Imaging*, 15:327–332, 1996.

[46] G. Dal Maso. *An introduction to Γ-convergence*, volume 8. Springer Science & Business Media, 2012.

[47] M. Dawood, F. Gigengack, X. Jiang, and K. P. Schäfers. A mass conservation-based optical flow method for cardiac motion correction in 3D-PET. *Med. Phys.*, 40(1):012505, 2013.

[48] A. Dax. On row relaxation methods for large constrained least squares problems. *SIAM J. Sci. Comput.*, 14(3):570–584, 1993.

[49] L. Desbat, S. Roux, and P. Grangeat. Compensation of some time dependent deformations in tomography. *IEEE Trans. Med. Imaging*, 26:261–269, 2007.

[50] H. Dirks. *Variational methods for joint motion estimation and image reconstruction.* PhD thesis, Universitäts-und Landesbibliothek Münster, 2015.

[51] H. Dirks. Joint large-scale motion estimation and image reconstruction. *arXiv preprint arXiv:1610.09908*, 2016.

[52] S. Dittmer, T. Kluth, M. T. R. Henriksen, and P. Maass. Deep image prior for 3D magnetic particle imaging: A quantitative comparison of regularization techniques on Open MPI dataset. *Int. J. Magn. Part. Imaging*, 7(1), 2021.

[53] M. J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. S. Duncan, L. Pizarro, D. Atkinson, B. F. Hutton, S. Ourselin, K. Thielemans, and S. R. Arridge. PET reconstruction with an anatomical MRI prior using parallel level sets. *IEEE Trans. Med. Imaging*, 35(9):2189–2199, 2016.

[54] I. Ekeland and R. Temam. *Convex analysis and variational problems.* SIAM, 1999.

[55] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[56] W. Erb, A. Weinmann, M. Ahlborg, C. Brandt, G. Bringout, T. M. Buzug, J. Frikel, C. Kaethner, T. Knopp, T. März, M. Möddel, M. Storath, and A. Weber. Mathematical analysis of the 1D model and reconstruction schemes for magnetic particle imaging. *Inverse probl.*, 34(5):055012, 2018.

[57] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015–1046, 2010.

[58] A. Fasano and A. Sequeira. *Hemomath: The mathematics of blood*, volume 18. Springer, 2017.

[59] D. Fortun, P. Bouthemy, and C. Kervrann. Optical flow modeling and computation: A survey. *Comput. Vis. Image Underst.*, 134:1–21, 2015.

[60] J. Franke, R. Lacroix, H. Lehr, M. Heidenreich, U. Heinen, and V. Schulz. MPI flow analysis toolbox exploiting pulsed tracer information – an aneurysm phantom proof. *Int. J. Mag. Part. Imag.*, 3(1), 2017.

[61] L. Frerking. *Variational methods for direct and indirect tracking in dynamic imaging*. PhD thesis, Westfälische Wilhelms-Universität Münster, 2016.

[62] B. Galvin, B. McCane, K. Novins, D. Mason, and S. Mills. Recovering motion fields: An analysis of eight optical flow algorithms. In *Proc. 1998 British Machine Vision Conference, BMVC*, volume 98, 1998.

[63] N. Gdaniec, M. Boberg, M. Möddel, P. Szwargulski, and T. Knopp. Suppression of motion artifacts caused by temporally recurring tracer distributions in multi-patch magnetic particle imaging. *IEEE Trans. Med. Imaging*, 39(11):3548–3558, 2020.

[64] N. Gdaniec, M. Schlüter, M. Möddel, M. G. Kaul, K. Krishnan, A. Schlaefer, and T. Knopp. Detection and compensation of periodic motion in magnetic particle imaging. *IEEE Trans. Med. Imaging*, 36:1511–1521, 2017.

[65] D. Gilland, B. Mair, J. Bowsher, and R. Jaszczak. Simultaneous reconstruction and motion estimation for gated cardiac ECT. *IEEE Trans. Nucl.*, 49(5):2344–2349, 2002.

[66] B. Gleich and J. Weizenecker. Tomographic imaging using the nonlinear response of magnetic particles. *Nature*, 435:1214–1217, 2005.

[67] M. Gräser, T. Knopp, P. Szwargulski, T. Friedrich, A. von Gladiss, M. Kaul, K. M. Krishnan, H. Ittrich, G. Adam, and T. M. Buzug. Towards picogram detection of superparamagnetic iron-oxide particles using a gradiometric receive coil. *Sci. Rep.*, 7(1):1–13, 2017.

[68] A. Güngör, B. Askin, D. A. Soydan, C. B. Top, E. U. Saritas, and T. Çukur. DEQ-MPI: A deep equilibrium reconstruction with learned consistency for magnetic particle imaging. *IEEE Trans. Med. Imaging*, 43(1):321–334, 2023.

[69] J. Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*, volume 15. Yale university press, 1923.

[70] J. Haegele, J. Rahmer, B. Gleich, J. Borgert, H. Wojtczyk, N. Panagiotopoulos, T. Buzug, J. Barkhausen, and F. Vogt. Magnetic particle imaging: visualization of instruments for cardiovascular intervention. *Radiology*, 265:933–938, 2012.

[71] B. N. Hahn. Efficient algorithms for linear dynamic inverse problems with known motion. *Inverse Probl.*, 30:035008, 20, 2014.

[72] B. N. Hahn. Motion estimation and compensation strategies in dynamic computerized tomography. *Sens. Imaging*, 18:1–20, 2017.

[73] A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu, and J. A. Steeden. Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease. *Magn. Reson. Med.*, 81(2):1143–1156, 2019.

[74] A. Hauptmann, O. Öktem, and C. Schönlieb. Image reconstruction in dynamic inverse problems with temporal models. In *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pages 1–31. Springer, 2021.

[75] B. Hofmann, B. Kaltenbacher, C. Poeschl, and O. Scherzer. A convergence rates result for tikhonov regularization in banach spaces with non-smooth operators. *Inverse Probl.*, 23(3):987, 2007.

[76] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artif. Intell.*, 17(1-3):185–203, 1981.

[77] S. Ilbey, C. B. Top, T. Çukur, E. U. Saritas, and H. E. Güven. Image reconstruction for magnetic particle imaging using an augmented lagrangian method. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 64–67. IEEE, 2017.

[78] A. Isola, A. Ziegler, T. Koehler, W. Niessen, and M. Grass. Motion-compensated iterative cone-beam CT image reconstruction with adapted blobs as basis functions. *Phys. Med. Biol.*, 53:6777–6797, 2008.

[79] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Int. Acad. Pol. Sei. Lett. A*, pages 335–37, 1937.

[80] B. Kaltenbacher. All-at-once versus reduced iterative methods for time dependent inverse problems. *Inverse Probl.*, 33:064002, 2017.

[81] B. Kaltenbacher, T. Schuster, and A. Wald. *Time-Dependent Problems in Imaging and Parameter Identification.* Springer, 2021.

[82] A. Katsevich. An accurate approximate algorithm for motion compensation in two-dimensional tomography. *Inverse Probl.*, 26:065007, 2010.

[83] A. Katsevich, M. Silver, and A. Zamayatin. Local tomography and the motion estimation problem. *SIAM J. Imaging Sci.*, 4:200–219, 2011.

[84] M. G. Kaul, J. Salamon, T. Knopp, H. Ittrich, G. Adam, H. Weller, and C. Jung. Magnetic particle imaging for in vivo blood flow velocity measurements in mice. *Phys. Med. Biol.*, 63(6):064001, 2018.

[85] S. Kindermann and A. Leitão. On regularization methods for inverse problems of dynamic type. *Num. Funct. Anal. Optim.*, 27:139–160, 2006.

[86] A. Kirsch et al. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer, 2011.

[87] T. Kluth. Mathematical models for magnetic particle imaging. *Inverse Probl.*, 34(8):083001, 2018.

[88] T. Kluth, B. Hahn, and C. Brandt. Spatio-temporal concentration reconstruction using motion priors in magnetic particle imaging. In T. Knopp and T. M. Buzug, editors, *9th International Workshop on Magnetic Particle Imaging 2019*, pages 23–24. Infinite Science Publishing, 2019.

[89] T. Kluth and B. Jin. L1 data fitting for robust reconstruction in magnetic particle imaging: quantitative evaluation on Open MPI dataset. *Int. J. Magn. Part. Imaging*, 6(2), 2020.

[90] T. Kluth, B. Jin, and G. Li. On the degree of ill-posedness of multi-dimensional magnetic particle imaging. *Inverse Probl.*, 34(9):095006, 2018.

[91] T. Knopp and T. M. Buzug. *Magnetic particle imaging: an introduction to imaging principles and scanner instrumentation.* Springer Science & Business Media, 2012.

[92] T. Knopp, P. Jürß, and M. Grosser. A deep learning approach for automatic image reconstruction in MPI. *Int. J. Magn. Part. Imaging*, 9(1, suppl. 1), 2023.

[93] T. Knopp, J. Rahmer, T. F. Sattel, S. Biederer, J. Weizenecker, B. Gleich, J. Borgert, and T. M. Buzug. Weighted iterative reconstruction for magnetic particle imaging. *Phys. Med. Biol.*, 55(6):1577, 2010.

[94] T. Knopp, P. Szwargulski, F. Griese, and M. Gräser. OpenMPIData: An initiative for freely accessible magnetic particle imaging data. *Data in brief*, 28:104971, 2020.

[95] T. Knopp, K. Them, M. Kaul, and N. Gdaniec. Joint reconstruction of non-overlapping magnetic particle imaging focus-field data. *Phys. Med. Biol.*, 60(8):L15, 2015.

[96] A. Kofler, M. Dewey, T. Schaeffter, C. Wald, and C. Kolbitsch. Spatio-temporal deep learning-based undersampling artefact reduction for 2D radial cine MRI with limited training data. *IEEE Trans. Med. Imaging*, 39(3):703–717, 2019.

[97] R. Kress. *Linear Integral Equations*, volume 3. Springer New York, 2014.

[98] J. Lampe, C. Bassoy, J. Rahmer, J. Weizenecker, H. Voss, B. Gleich, and J. Borgert. Fast reconstruction in magnetic particle imaging. *Phys. Med. Biol.*, 57(4):1113, 2012.

[99] L. F. Lang, S. Neumayer, O. Öktem, and C.-B. Schönlieb. Template-based image reconstruction from sparse tomographic data. *Appl. Math. Optim.*, 82:1081–1109, 2020.

[100] T. Li, M. Zhang, W. Qi, E. Asma, and J. Qi. Deep learning based joint PET image reconstruction and motion estimation. *IEEE Trans. Med. Imaging*, 41(5):1230–1241, 2022.

[101] F. Lieb and T. Knopp. A sparse row-action algorithm for magnetic particle imaging. *Int. J. Magn. Part. Imaging*, 6(2 Suppl 1), 2020.

[102] F. Lieb and T. Knopp. A wavelet-based sparse row-action method for image reconstruction in magnetic particle imaging. *Med. Phys.*, 48(7):3893–3903, 2021.

[103] J.-L. Lions. Quelques méthodes de résolution de problemes aux limites non linéaires. 1969.

[104] J. Liu, X. Zhang, X. Zhang, H. Zhao, Y. Gao, D. Thomas, D. Low, and H. Gao. 5D respiratory motion model based image reconstruction algorithm for 4D cone-beam computed tomography. *Inverse Probl.*, 31:115007, 2015.

[105] W. Lu and T. R. Mackie. Tomographic motion detection and correction directly in sinogram space. *Phys. Med. Biol.*, 47(8):1267–1284, 2002.

[106] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. 7th Int. Joint Conf. Artif. Intell. (IJCAI)*, pages 121–130, 1981.

[107] F. Lucka, N. Huynh, M. Betcke, E. Zhang, P. Beard, B. Cox, and S. Arridge. Enhancing compressed sensing 4D photoacoustic tomography by simultaneous motion estimation. *SIAM J. Imaging Sci.*, 11(4):2224–2253, 2018.

[108] P. Ludewig, N. Gdaniec, J. Sedlacik, N. D. Forkert, P. Szwargulski, M. Graeser, G. Adam, M. G. Kaul, K. M. Krishnan, R. M. Ferguson, A. P. Khandhar, P. Walczak, J. Fiehler, G. Thomalla, C. Gerloff, and T. Knopp. Magnetic particle imaging for real-time perfusion imaging in acute stroke. *ACS Nano.*, 11(10):10480–10488, 2017.

[109] P. Ludewig, M. Graeser, N. D. Forkert, F. Thieben, J. Rández-Garbayo, J. Rieckhoff, K. Lessmann, F. Förger, P. Szwargulski, T. Magnus, and T. Knopp. Magnetic particle imaging for assessment of cerebral perfusion and ischemia. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.*, 14(1):e1757, 2022.

[110] M. Maass, C. Droigk, F. Katzberg, P. Koch, and A. Mertins. A recovery algorithm based on the kaczmarz algorithm and ADMM splitting with application to convex optimization in magnetic particle imaging. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 2135–2139. IEEE, 2021.

[111] E. Meinhardt-Llopis, J. S. Pérez, and D. Kondermann. Horn-schunck optical flow with a multi-scale strategy. *Image Processing On Line*, 2013:151–172, 2013.

[112] J. L. Mueller and S. Siltanen. *Linear and nonlinear inverse problems with practical applications*. SIAM, 2012.

[113] M. Negahdar and A. A. Amini. Planar deformable motion estimation incorporating mass conservation and image gradient constancy. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 536–540. IEEE, 2011.

[114] M. Nitzsche and B. N. Hahn. Dynamic image reconstruction in MPI with RESESOP-Kaczmarz. *Int. J. Mag. Part. Imag.*, 10(1), 2024.

[115] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the Mumford-Shah functional. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1133–1140. IEEE, 2009.

[116] J. Rahmer, J. Weizenecker, B. Gleich, and J. Borgert. Analysis of a 3-D system function measured for magnetic particle imaging. *IEEE Trans. Med. Imaging*, 31(6):1289–1299, 2012.

[117] M. Reyes, G. Malandain, P. Koulibaly, M. González-Ballester, and J. Darcourt. Model-based respiratory motion compensation for emission tomography image reconstruction. *Phys. Med. Biol.*, 52:3579–3600, 2007.

[118] A. Rieder. *Keine Probleme mit inversen Problemen: eine Einführung in ihre stabile Lösung*. Springer-Verlag, 2013.

[119] R. Rockafellar. Convex analysis. *Princeton Math. Series*, 28, 1970.

[120] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

[121] S. Roux, L. Desbat, A. Koenig, and P. Grangeat. Exact reconstruction in 2D dynamic CT: compensation of time-dependent affine deformations. *Phys. Med. Biol.*, 49(11):2169, 2004.

[122] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D: Nonlinear Phenom.*, 60(1-4):259–268, 1992.

[123] E. U. Saritas, P. W. Goodwill, G. Z. Zhang, and S. M. Conolly. Magnetostimulation limits in magnetic particle imaging. *IEEE Trans. Med. Imaging*, 32(9):1600–1610, 2013.

[124] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 1. Springer, 2009.

[125] I. Schmale, B. Gleich, J. Rahmer, C. Bontus, J. Schmidt, and J. Borgert. MPI safety in the view of MRI safety standards. *IEEE Trans. Magn.*, 51(2):1–4, 2015.

[126] U. Schmitt and A. Louis. Efficient algorithms for the regularization of dynamic inverse problems: I. Theory. *Inverse Probl.*, 18:645–658, 2002.

[127] U. Schmitt, A. Louis, C. Wolters, and M. Vauhkonen. Efficient algorithms for the regularization of dynamic inverse problems: II. Applications. *Inverse Probl.*, 18:659–676, 2002.

[128] C. Schnörr. Determining optical flow for irregular domains by minimizing quadratic functionals of a certain class. *Int. J. Comput. Vis.*, 6(1):25–38, 1991.

[129] T. Schuster, B. Hahn, and M. Burger. Dynamic inverse problems: modelling—regularization—numerics. *Inverse Probl.*, 34(4):040301, 2018.

[130] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization methods in Banach spaces*, volume 10. Walter de Gruyter, 2012.

[131] J. Simon. Compact sets in the space $L^p(0,T;B)$. *Annali di Matematica pura ed applicata*, 146:65–96, 1986.

[132] H. Spies, H. Haußecker, B. Jähne, and J. L. Barron. Differential range flow estimation. In *Mustererkennung 1999*, pages 309–316. Springer, 1999.

[133] H. Spies, B. Jähne, and J. L. Barron. Range flow estimation. *Comput. Vis. Image Underst.*, 85(3):209–231, 2002.

[134] M. Storath, C. Brandt, M. Hofmann, T. Knopp, J. Salamon, A. Weber, and A. Weinmann. Edge preserving and noise reducing reconstruction for magnetic particle imaging. *IEEE Trans. Med. Imaging*, 36(1):74–85, 2016.

[135] N. G. T. Knopp and M. Möddel. Magnetic particle imaging: From proof of principle to preclinical applications. *Phys. Med. Biol.*, 62(14):R124, 2017.

[136] K. Them, M. G. Kaul, C. Jung, M. Hofmann, T. Mummert, F. Werner, and T. Knopp. Sensitivity enhancement in magnetic particle imaging by background subtraction. *IEEE Trans. Med. Imaging*, 35(3):893–900, 2015.

[137] K. Them, J. Salamon, P. Szwargulski, S. Sequeira, M. Kaul, C. Lange, H. Ittrich, and T. Knopp. Increasing the sensitivity for stem cell monitoring in system-function based magnetic particle imaging. *Phys. Med. Biol.*, 61(9):3279, 2016.

[138] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. Comput. Vis.*, 9:137–154, 1992.

[139] A. Tsanda, P. Jürß, N. Hackelberg, M. Grosser, M. Möddel, and T. Knopp. Extension of the kaczmarz algorithm with a deep plug-and-play regularizer. *Int. J. Magn. Part. Imaging*, 10(1, suppl. 1), 2024.

[140] J. Weickert, A. Bruhn, T. Brox, and N. Papenberg. A survey on variational optic flow methods for small displacements. In O. Scherzer, editor, *Mathematical models for registration and applications to medical imaging*, Mathematics in industry, pages 103–136. Springer, Berlin, Heidelberg, 2006.

[141] J. Weizenecker, B. Gleich, and J. Borgert. Magnetic particle imaging using a field free line. *J. Phys. D Appl. Phys.*, 41(10):105009, 2008.

[142] J. Weizenecker, B. Gleich, J. Rahmer, H. Dahnke, and J. Borgert. Three-dimensional real-time in vivo magnetic particle imaging. *Phys. Med. Biol.*, 54:L1, 2009.

[143] E. Y. Yu, M. Bishop, B. Zheng, R. M. Ferguson, A. P. Khandhar, S. J. Kemp, K. M. Krishnan, P. W. Goodwill, and S. M. Conolly. Magnetic particle imaging: a novel in vivo imaging platform for cancer detection. *Nano letters*, 17(3):1648–1654, 2017.

[144] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29*, pages 214–223. Springer, 2007.

[145] L. Zdun, M. Boberg, and C. Brandt. Joint multi-patch reconstruction: fast and improved results by stochastic optimization. *Int. J. Magn. Part. Imaging*, 8(2), 2022.

[146] L. Zdun and C. Brandt. Fast MPI reconstruction with non-smooth priors by stochastic optimization and data-driven splitting. *Phys. Med. Biol.*, 66(17):175004, 2021.

[147] B. Zheng, T. Vazin, P. W. Goodwill, A. Conway, A. Verma, E. Ulku Saritas, D. Schaffer, and S. M. Conolly. Magnetic particle imaging tracks the long-term fate of in vivo neural cell implants with high image contrast. *Sci. Rep.*, 5(1):14055, 2015.

[148] T. Zhu, L. Yin, J. He, Z. Wei, X. Yang, J. Tian, and H. Hui. Accurate concentration recovery for quantitative magnetic particle imaging reconstruction via nonconvex regularization. *IEEE Trans. Med. Imaging*, 2024.

List of Algorithms

List of Figures

List of Tables

A. Details and Additional Results for the Numerical Experiments

A.1. Simulated data

This section reveals detailed information about the simulation setting and the parameters used for reconstruction of the simulated MPI data. We first describe the simulated scanner in Section A.1.1 and then comment on the search areas and optimal reconstruction parameters in Section A.1.2.

A.1.1. Properties of the modeled MPI scanner

The scanner setup used for simulations is chosen in order to match the Bruker Biospin scanner (Ettlingen) at the University Medical Center Hamburg-Eppendorf. This simulated scanner is used to produce the data used in Section 3.3, Section 6.1 and Section 6.4. The drive-field frequencies are defined by the base frequency 2.5 MHz and different frequency dividers, i.e., $f_x = 2.5\,\mathrm{MHz}/102$, $f_y = 2.5\,\mathrm{MHz}/96$ and $f_z = 2.5\,\mathrm{MHz}/99$. This yields an amplitude of $14\,\mathrm{mT}/\mu_0$ and a repetition time of 21.54 ms. The selection field gradient strength is $0.5\,\mathrm{T/m}/\mu_0$ in x- and y-direction and $1.0\,\mathrm{T/m}/\mu_0$ in z-direction. The magnetic nanoparticles are modeled as stated in Table A.1.

Table A.1.: Physical constants and parameters for modeling the magnetic nanoparticles.

Parameter	Value
Permeability constant μ_0	$4\pi \cdot 10^{-7}\mathrm{N/A}^2$
Particle core diameter	$2 \cdot 10^{-8}\mathrm{m}$
Core saturation magnetization	$0.6\,\mathrm{T}/\mu_0$

We model ideal magnetic fields and use the Langevin magnetization model. In the forward operator voltage measurements for three receive channels in respective unit vector directions are used. To avoid inverse crime, simulations are performed on a finer grid compared to the reconstruction, i.e., we use 40^3 voxels of size $1\,\mathrm{mm} \times 1\,\mathrm{mm} \times 0.5\,\mathrm{mm}$ for simulation and 20^3 voxels of size $2\,\mathrm{mm} \times 2\,\mathrm{mm} \times 1\,\mathrm{mm}$ for reconstruction. For reconstruction, the time-dependent signals are transformed to the Fourier domain, a frequency selection is performed and real and imaginary part are split, i.e., concatenated. In the subframe reconstruction setting, the data are transformed back to the time domain after frequency selection.

A.1.2. Parameter search region and optimal parameters

For the static image reconstruction task in Section 3.3 we summarize the parameter search region in Table A.2 and the optimal parameters in Table A.3. Optimality is considered with respect to the SSIM value reached by reconstructed images with the respective parameter combination.

Table A.2.: Parameter search area for static image reconstruction for simulated data.

Algorithm	Par. 1	Min	Max	Par. 2	Min	Max
Kac	λ	10^{-6}	10^{3}	k	1	20
FL-L1D	α	10^{-8}	10^{+1}	β	10^{-8}	10^{-1}
FL-L2D	α	10^{-8}	10^{+0}	β	10^{-8}	10^{+0}
L1-L1D				β	10^{-8}	10^{+1}
L2Grad-L2D	α	10^{-8}	10^{+1}			

Table A.3.: Optimal parameters with respect to the SSIM value for static image reconstruction for simulated data.

	"Lemon"		"Dots"		"Pi"	
Algorithm	Par. 1	Par. 2	Par. 1	Par. 2	Par. 1	Par. 2
Kac	$\lambda = 10^{-1}$	$k = 18$	$\lambda = 10^{-5}$	$k = 7$	$\lambda = 10^{+0}$	$k = 9$
FL-L1D	$\alpha = 10^{-1}$	$\beta = 10^{-2}$	$\alpha = 10^{+0}$	$\beta = 10^{-2}$	$\alpha = 10^{+0}$	$\beta = 10^{-3}$
FL-L2D	$\alpha = 10^{-1}$	$\beta = 10^{-1}$	$\alpha = 10^{-2}$	$\beta = 10^{-1}$	$\alpha = 10^{-1}$	$\beta = 10^{-3}$
L1-L1D		$\beta = 10^{-1}$		$\beta = 10^{-3}$		$\beta = 10^{+0}$
L2Grad-L2D	$\alpha = 10^{-2}$		$\alpha = 10^{+0}$		$\alpha = 10^{-2}$	

For the dynamic simulated data in Section 6.1, we used the search regions described by Table A.4 for the joint approaches. Parameter tests for the joint approaches were performed on 10 time steps simultaneously. The resulting image sequences for each parameter

Table A.4.: Parameter search area for the joint approaches for the simulated data.

Parameter	Min. value	Max. value
α_1	10^{-3}	10^{-1}
α_2	10^{-8}	10^{-5}
β	10^{-2}	1
γ	10^{-5}	1

combination was tested for the SSIM value. The best parameter combination is stated in Table A.5.

For the Kaczmarz algorithm, we tested for the early stopping index k and the Tikhonov regularization parameter λ. The parameter λ was tested in the range between 10^{-5} and 10^{3}, k in between 1 and 10.

Frame-by-frame SPDHG algorithm has two parameters, α_1 corresponding to the L^1-penalty term and α_2 corresponding to the TV penalty term. The testing range for α_1 was in between 10^{-3} and 10^{-1}, for α_2 in between 10^{-9} and 10^{-5}.

Table A.5.: Optimal parameters with respect to the SSIM value for joint image reconstruction and motion estimation for simulated data.

Algorithm	Motion model	Early stopping	λ	α_1	α_2	β	γ
Kaczmarz	None	$k = 3$	10^3				
SPDHG	None	$k = 2 \cdot 10^3$		10^{-2}	10^{-7}		
L1OF	Optical Flow	$k = 10^2$		10^{-2}	10^{-7}	10^{-1}	10^{-4}
L2OF	Optical Flow	$k = 10^2$		10^{-2}	10^{-7}	10^{-1}	10^{-4}
L1MC	Mass Conservation	$k = 10^2$		10^{-2}	10^{-8}	10^{-1}	10^{-4}
L2MC	Mass Conservation	$k = 10^2$		10^{-2}	10^{-5}	10^{-1}	10^{-4}

A.2. Rotation phantom data

A.2.1. Parameter search region and optimal parameters

For the Kaczmarz algorithm, we tested for the early stopping index k and the Tikhonov regularization parameter λ. The parameter λ was tested in the range between 10^{-4} and 30, k in between 1 and 100.

Frame-by-frame SPDHG algorithm has two parameters, α_1 corresponding to the L^1-penalty term and α_2 corresponding to the TV penalty term. The testing range for α_1 was in between 10^{-4} and 1, for α_2 in between 10^{-4} and 10.

Parameter tests for the joint approaches were performed on 15 time steps simultaneously within the ranges stated in Table A.6.

Table A.6.: Parameter search area for the joint approaches for the rotation phantom data.

Parameter	Min. value	Max. value
α_1	10^{-2}	10
α_2	10^{-2}	100
β	10^{-2}	1
γ	1	200

The resulting image sequences were visually inspected and the most convincing ones were chosen. The search area was iteratively reduced and refined until no visual differences were observed.

The resulting parameters for the 3 Hz dataset are stated in Table 6.4. Parameter tests for the 1 Hz and 7 Hz datasets are performed analogously and yield similar results.

A.2.2. Additional illustrative results for the 7 Hz dataset

Considering the 7 Hz data set, we obtain the trajectories depicted in Figure A.1 from the motion estimates. In the left column, we depict the optimal trajectory constructed manually from the reconstructed image sequences. In the middle, we have the piecewise linear trajectories obtained by adding up motion estimates at different time steps. In the right column, we have a smoothed version of the curve in the middle. The yellow circle with correct diameter is fitted for guidance. The trajectories obtained based on the image sequences computed by Kaczmarz method, L2OF and Joint reconstruction algorithm using an L^1-data discrepancy term and the mass conservation motion model (L1MC) severely underestimate the magnitude of motion.

In Table A.7, we state the computed centers (y, z) and radii r of the best-fit circles fitted to the coordinates obtained by the piecewise linear trajectories in the middle column of Figure A.1. The computed values confirm the visual impression, the radii of the Kaczmarz method, L2OF and L1MC are too small. The reference position for the center of the yellow circles is $(y_{\text{ref}}, z_{\text{ref}}) = (12.5, 13)$. Please note that the position of the reference is not exact but only a guess based on the reconstructed images. The center of the L1OF is close in z-direction but not in y-direction, the center of the L2MC is close in y- but not in z-direction. Those two methods yield the best overall results.

Table A.7.: Centers (y, z) and radii r of the best-fit circles fitted to the coordinates obtained by following the motion estimates on each time step for the 7 Hz dataset. The reference value is an approximate value based on visual inspection of the image sequences combined with knowledge from the experimental setup.

	Reference	Kaczmarz	SPDHG	L1OF	L2OF	L1MC	L2MC
y	12.5	11.81	11.16	13.91	11.88	12.50	12.22
z	13	15.11	14.23	13.23	15.93	14.00	11.35
r	5.75	2.80	3.45	5.45	4.10	4.17	5.35

A.3. Mouse data

A.3.1. Parameter search region and optimal parameters

Motivated by the specific application, we use the L1MC algorithm. The parameter search area is indicated in Table A.8, tests were performed on 30 time steps simultaneously.

Table A.8.: Parameter search area for reconstruction of in-vivo mouse data by the L1MC algorithm.

Parameter	Min. value	Max. value
α_1	10^{-2}	10^{-1}
α_2	10^{-3}	200
β	10^{-3}	10^{-1}
γ	1	1000

The resulting reconstruction parameters are given by $\alpha_1 = 0.05$, $\alpha_2 = 100$, $\beta = 0.1$ and $\gamma = 100$. They are chosen based on visual inspection of the resulting image sequences.

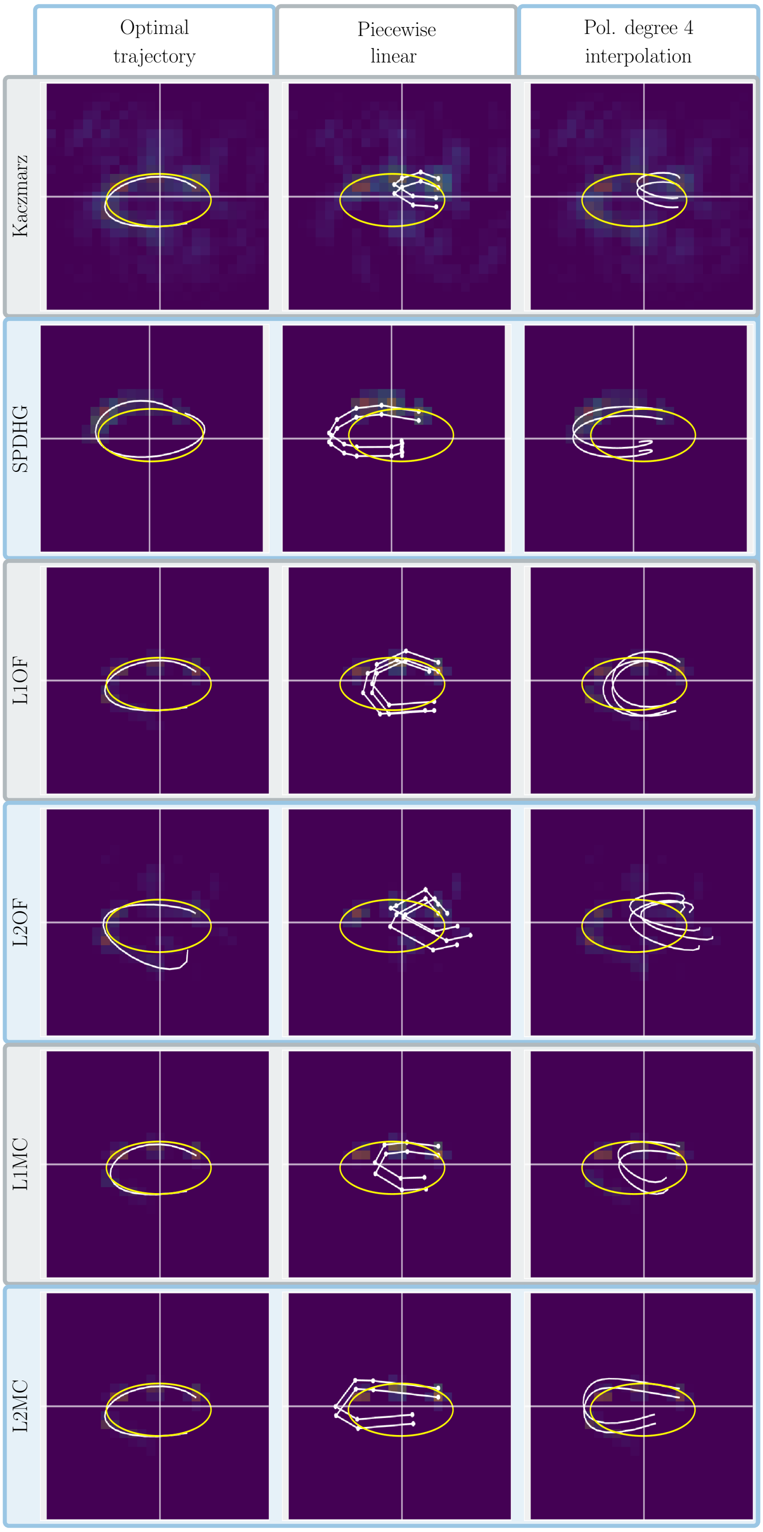


Figure A.1.: Trajectories of dynamic particle concentration based on motion estimates by various algorithmic approaches for the 7 Hz rotation phantom dataset. The left column displays the optimal particle trajectory as observed in the reconstructed image sequence for the 7 Hz dataset by following the maximum concentration value along the frames and smoothing the result by a 4th order polynomial approximation in white. For guidance, we fit a circle with correct diameter to the curve in yellow. The middle column shows the piecewise linear trajectories which are obtained by adding up the computed motion of every time step. The right column displays a 4th order polynomial approximation to the curve in the middle. The trajectories based on SPDHG, L1OF and L2MC match the correct path best.

B. A Result on the Decay of Singular Values in MPI

B.1. Theorem 4.1. in [90]

In this section, we state the setting and result of [90, Theorem 4.1.] using our notation. An offspring of the proof is given by the fact, that the system function of MPI lies in the space $H^0\left(I; L^2(\Omega)\right)$. This is used for the proof of Lemma 3.1 and we thus state the original theorem and proof here for completeness.
The static MPI forward operator is defined by $A : L^2(\Omega) \to L^2(\Omega)$

$$
\begin{aligned}
c &\mapsto \int_\Omega c(x)s(x,t)\mathrm{d}x, \\
s &= \mu_0 m_0 R^T \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\mathcal{L}_\beta\left(\|H\|\right)}{\|H\|}H\right], \\
H\left(x,t\right) &= H_D(t) + H_S(x).
\end{aligned}
$$

Theorem B.1: Theorem 4.1. in [90]

Let $0 < \beta < \infty$, $d = 1,2,3$, $-H_D \in \left(H^s(I)\right)^d$ with $s \geq 1$, $H_s \in \left(L^\infty(\Omega)\right)^d$ and $R \in \left(L^\infty(\Omega)\right)^d$.
Then for the operator $A : L^2(\Omega) \to L^2(\Omega)$ defined above, the singular values σ_n decay as $\sigma_n \leq C n^{\frac{1}{2}-s}$.

Proof. Since $-H_D \in \left(H^s(I)\right)^d$ and $H_s \in \left(L^\infty(\Omega)\right)^d$, it holds for the effective magnetic field and its derivative that $H\left(x,t\right) = H_D(t) + H_S(x) \in H^s\left(I; \left(L^\infty(\Omega)\right)^d\right) \subset L^\infty\left(I; \left(L^\infty(\Omega)\right)^d\right)$ and $\dot{H}(x,t) = \dot{H}_D(t) \in \left(H^{s-1}(I)\right)^d$ by Sobolev embedding. Considering $R \in \left(L^\infty(\Omega)\right)^d$, we note that

$$
\begin{aligned}
R^T H &\in H^s\Big(I; \left(L^\infty(\Omega)\right)\Big), \\
R^T \dot{H} &\in \Big(H^{s-1}\big(I; L^\infty(\Omega)\big)\Big).
\end{aligned}
$$

By [90, Lemma 2.1.], we further know that $\frac{\mathcal{L}_\beta(\sqrt{z})}{\sqrt{z}} \in C_b^\infty([0,\infty))$. This property is proven directly in the lemma. Moreover, $\|H^2\| \in H^s\big(I; L^\infty(\Omega)\big)$ for $s \geq 1$ (cf. [90, Theorem 3.1.]). The theorem considers pointwise multiplication on Sobolev spaces in general. With the help of [90, Lemma 3.1.(i)], a lemma on composition operators on Sobolev spaces, it can be deduced that

$$
\begin{aligned}
\frac{\mathcal{L}_\beta(\|H\|)}{\|H\|} &\in H^s\big(I; L^\infty(\Omega)\big), \\
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathcal{L}_\beta(\|H\|)}{\|H\|} &\in H^{s-1}\big(I; L^\infty(\Omega)\big).
\end{aligned}
$$

It thus follows that

$$
\begin{aligned}
s &= \mu_0 m_0 R^T \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\mathcal{L}_\beta\left(\|H\|\right)}{\|H\|} H\right] \\
&= \mu_0 m_0 \left[R^T H \frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathcal{L}_\beta\left(\|H\|\right)}{\|H\|} + \frac{\mathcal{L}_\beta\left(\|H\|\right)}{\|H\|} R^T \dot{H}\right] \\
&\in H^{s-1}\big(I; L^\infty(\Omega)\big) \subset H^{s-1}\left(I; L^2(\Omega)\right).
\end{aligned}
$$

The assertion about the singular value decay then follows from [90, Theorem 3.2.]. $\square$

Declaration of Authorship/ Eidesstattliche Erklärung

Hiermit erkläre ich, Lena Westen, an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ich versichere außerdem, dass die elektronische Version dieser Dissertation mit der gebundenen, bei der Fakultät zur Archivierung eingereichten Version übereinstimmt.

I, Lena Westen, hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Furthermore, I declare that the electronic version of this dissertation coincides with the printed bound copy submitted to the faculty for archiving.

City, Date | Ort, Datum   Signature | Unterschrift

List of Publications

During the doctorate, the candidate has contributed to the following research publications:

[27] C. Brandt, T. Kluth, T. Knopp and L. Westen. Dynamic image reconstruction with motion priors in application to three dimensional magnetic particle imaging. *SIAM J. Imaging Sci.*, 17(3):1539-1586, 2024.
[145] L. Zdun, M. Boberg and C. Brandt. Joint multi-patch reconstruction: fast and improved results by stochastic optimization. *Int. J. Magn. Part. Imaging*, 8(2), 2022.
[146] L. Zdun and C. Brandt. Fast MPI reconstruction with non-smooth priors by stochastic optimization and data-driven splitting. *Phys. Med. Biol.*, 66(17):175004, 2021.

Declaration of Contributions

This declaration describes which parts of the thesis were obtained with contributions of other people and what their respective contributions are. Parts of Sections 3.1, 3.2, 4.1, 4.2, 5.1 and 6.1 to 6.3 have been first published in [27] and have been prepared in collaboration with the co-authors.
More precisely, the idea and conceptualization of [27] are equally contributed by C. Brandt, T. Kluth and L. Westen. In Chapter 3, Lemma 3.1 and 3.2 as well as Theorem 3.1 are equally due to T. Kluth and L. Westen. Algorithmic details and implementations as described in Sections 4.1 and 4.2 are mainly due to L. Westen. The proof of Theorem 5.1 in Section 5.1 is mainly due to L. Westen with minor contributions by T. Kluth. The numerical experiments in Sections 6.1 to 6.3 have been planned by all four authors and have been carried out mainly by L. Westen with minor contributions by C. Brandt.