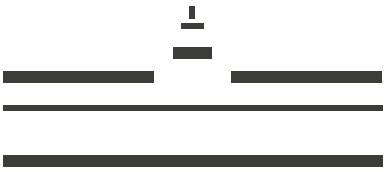

Westfälische
Wilhelms-Universität
Münster

Fach: Mathematik

Variational Methods for Direct and Indirect Tracking in Dynamic Imaging

Inaugural Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften
– Dr. rer. nat. –
im Fachbereich Mathematik und Informatik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

eingereicht von

Lena Christina Frerking
aus Münster

– 2016 –

Dekan:	Prof. Dr. Xiaoyi Jiang
Erster Gutachter:	Prof. Dr. Martin Burger (WWU Münster)
Zweiter Gutachter:	Dr. Christoph Brune (University of Twente)
Tag der mündlichen Prüfung:	20.01.2017
Tag der Promotion:	20.01.2017

Abstract

The topic of tracking in terms of estimating motion is a major issue for a wide range of applications. In this thesis we study techniques for solving this task given different external requirements depending on the type of data. The thesis can roughly be divided into two theoretical and two applied parts.

In the first theoretical part we study models for direct tracking where no previous reconstruction is necessary. We extend established optical flow models in order to improve their performance in connection with complex motion, like large-scale movement or the combination of smooth regions and sharp edges in a single flow field. We analyze these models and check their performance based on different synthetic data sets. For the second theoretical part we consider types of data where tracking is not possible based on the raw data, but it requires reconstruction beforehand. We present a model for simultaneous reconstruction and motion estimation, which allows for a time-dependent reconstruction operator. Again, we analyze this model and discuss its numerical realization.

The applied parts consist of specific examples for both of the previously mentioned types of tracking. For the direct tracking model we study microscopy data of migrating cells. We illustrate the relevance of motion estimation for this data and explicate the required mathematical tasks. The application for indirect tracking comprises an approach for dynamic X-ray tomography. With the model for indirect tracking we show results for tomography of continuously moving objects while using only one or two X-ray sources.

Keywords: Image Processing, Inverse Problems, Variational Methods, Total Variation, Optical Flow, Joint Motion Estimation and Image Reconstruction, Microscopy Imaging, Time-Dependent Reconstruction, Dynamic X-Ray Tomography

Acknowledgments

I would like to thank

Martin Burger for giving me the opportunity to be a part of his work group and with that allowing me to gain so many valuable experiences, for supervising my thesis, for patiently answering all of my questions and for reliably responding to mails until late at night.

Christoph Brune for his supervision in the beginning of my time as a PhD student, for co-reviewing my thesis, and for all of the useful discussions about cells and optical flow.

my colleagues and former colleagues from the imaging work group and the whole Institute for Applied Mathematics in Münster, in particular

- *Hendrik Dirks* for providing his codes and helping me again and again to adapt them to my models, and for being a comfortable traveling partner for official trips.
- *Carolin Roßmanith* for proofreading, for being the best running mate ever who keeps me motivated even if its cold and rainy, and for always spreading optimism.
- *Eva-Maria Brinkmann* for proofreading, for always taking care of everyone and for never refusing any kind of help.
- *Sebastian Suhr* for proofreading, for always being one step ahead and generously sharing his gained experiences.
- *Daniel Tenbrinck* for proofreading, for sharing zebrafishes with me and for having useful advises concerning academic issues.
- *Pia Heins* for proofreading, for honestly sharing her experiences and having useful advices, and for the inspiring reports about her trip around the world.
- *Stefan Girke* for being incredibly patient while running with me even though being actually twice as fast as I am.

- *Joana Grah* for never shying away from any curious activities and thus providing extraordinary enjoyable times whenever we meet.
- *Andreas Nüßing* for being a great office mate, even if it was only for a year, and for countless enjoyable discussions during ice cream sessions.
- *Frank Wübbeling* for always being honest and straightforward and for organizing the skiing colloquium.
- *Claudia Giesbert* and *Carolin Gietz* for taking care of all the organizational things and for always having an open ear.

the people from the inverse problems group of the University of Helsinki for making my stay so enjoyable, especially

- *Samuli Siltanen* for his interest and all of the useful discussions about the shared project, for infecting everyone with his good temper, and for being a great conversationalist.
- *Tapio Helin* for inviting me to Helsinki and providing me the opportunity work on that interesting project.
- *Andreas Hauptmann* for showing me around in Helsinki and for his engagement in the shared project.

my parents *Christel Frerking* and *Norbert Frerking* for believing in me more than I do, for building me up when I am struggling and for supporting me in every possible way.

Simon Frerking for being my protective big brother and for always reminding me of the importance of solidarity and family.

Matthias Robbers for always being there for me and for standing by my side, for believing in me, for motivating me in so many ways and for being my calming counterbalance in times of doubt.

all of my friends for keeping me grounded and for always supporting me, in particular *Laura Oeing* and *Kirsten Rawert* for being the best friends I can imagine.

and everyone else who should be mentioned here, but has been forgotten. I am sorry for that!

Contents

List of Figures

List of Tables

1 Introduction

In this thesis we propose and investigate mathematical models for tracking, which are based on the estimation of flow fields. Furthermore, we present a model for indirect motion estimation via image reconstruction. The latter allows to indirectly track objects by using necessary reconstruction results. Before discussing these issues in detail, we start by giving the motivation for this work. This also includes a short introduction to the topic of inverse problems. After stating our contribution, we give an overview of organization of this work.

1.1 Motivation

Motion estimation is of great importance for many different areas. Biologists are e.g. interested in following single cells or analyzing their behavior over a couple of time steps. Physicians need motion estimation e.g. to analyze the blood flow. Meteorologists depend on motion estimation to predict the velocity and the direction of clouds and other weather situations. Also in the automotive sector, motion estimation gains more and more attention in recent years. With its help it is possible to automatically recognize people crossing the path of a car and to stop the car without any interaction of the driver. The whole issue of self-driving cars, which is already pretty far in development and will very likely be all around in the nearer future, is heavily depending on motion estimation.

Usually, motion estimation is based on two successive images taken e.g. out of a video sequence. It aims at answering the question of what happens between these images. The result is a transformation that, applied to a specific image, yields the subsequent image. The transformation consists of a vector field that assigns a direction and a

distance to every single pixel. Incorporating the time that passes between the recording of the two images, the distance of a vector is equivalent to a certain velocity.

1.2 Inverse Problems in Imaging

The issue of estimating motion belongs to the broad field of inverse problems. Generally, this means that we are confronted with a certain situation or effect, and are interested in the source or the cause of this situation. In the case of motion estimation, the situation is given by data in terms of the images at distinct time steps. The cause of the situation is the vector field, which transforms the first image into the second image. A generic inverse problem can often be written as

$$Av = \tilde{g},$$

where $\tilde{g}$ is the given measurement, $A : \mathcal{X} \to \mathcal{Y}$ is a linear and compact operator, and v is the cause that we are looking for. However, in image processing the effect, i.e. the data we are dealing with, is usually disturbed e.g by noise or other artifacts. This means that the data $\tilde{g}$ is composed of the actual data g and the mentioned artifacts σ. Therefore, the problem we need to solve is

$$Av = g + \sigma. \tag{1.1}$$

Solving the above problem with respect to u is usually an *ill-posed* problem. The differentiation between a *well-posed* and an *ill-posed* problem was already invented in 1902 by the french mathematician Jacques Hadamard. In Hadamard (1902) he proposed the following definition:

Definition 1.1. (Well-Posed Problem)
A problem is called *well-posed*, if

- there exists a solution of the problem,
- the solution is unique,
- the solution depends continuously on the input data.

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

In the case of (1.1), the compactness of the operator $\mathcal{A}$ usually prohibits its continuous inversion (cf. Engl et al. (1996)). Consequently, the third condition is violated and

therefore the problem is *ill-posed*. Even if we would be able to find an inverse $\mathcal{A}^{-1}$ of $\mathcal{A}$, it would still not be possible to calculate the exact solution due to the random appearance of the noise σ.

As a consequence of this ill-posedness, it is not possible to find an exact solution v of the problem. Thus, the goal is usually to find an approximate solution $\hat{v}$, which is close to the exact solution. A convenient way to find such an approximation is a variational approach. Here, the aim is to solve the following minimization problem:

$$\hat{v} = \underset{v \in dom A}{\arg\min} \{\mathcal{D}(Av, g) + \alpha \mathcal{R}(v)\}.$$

The first term $\mathcal{D}(Av, g)$ is the so-called *data fidelity*. It is usually a distance term between Av and g accomplished with a norm, i.e.

$$\mathcal{D}(Av, g) = \|Av - g\|,$$

where $\|\cdot\|$ is typically the L^1 norm or the squared L^2 norm. The minimization of this term ensures that the approximate solution $\hat{v}$ is close to the exact solution v. $\mathcal{R}(v)$ is referred to as the *regularizer*. It allows to incorporate a priori knowledge, which enhances specific properties of the solution. Common choices are e.g. certain derivatives of u again fitted with a norm. The parameter α is the *regularization parameter*. It serves as a weighting between the data fidelity and the regularizer.

1.3 Contributions

The idea to use the optical flow constraint in a variational framework first came up more than 30 years ago in Horn and Schunck (1981). Here, for both the data fidelity and the regularizer, the squared L^2 norm was used. In Aubert et al. (1999) the L^1 norm instead of the L^2 norm was introduced for the optical flow constraint. The basis of a nonlinear optical flow constraint was given in Papenberg et al. (2006). In this thesis we further discuss this topic and prove convergence of an iterative approach that approximates a nonlinear version of the optical flow constraint.

The usage of the total variation seminorm as a regularizer for optical flow was also discovered in Papenberg et al. (2006). The purpose of allowing smooth transitions and sharp edges was already discussed in Charbonnier et al. (1997) by the application of the Huber loss function as a regularizer. The idea was massively pushed forward with the invention of the total generalized variation in Bredies et al. (2010). In this thesis

we derive a general formulation of a regularizer, which can amongst others be reduced to the total generalized variation, as well as the Huber loss function. Additionally, we present a generalization of Bregman iterations, which can be applied to the nonlinear optical flow approach. Originally, the Bregman iterations were suggested in Osher et al. (2005) and so far mainly used for reconstruction purposes.

Combining motion estimation with image reconstruction in a variational setting was first introduced with gating approaches for cardiac imaging, like in Gilland et al. (2002). A more general type of motion was considered in Brune (2010), where the motion estimation takes place in terms of optimal transport. In recent years joint models were further studied in Suhr (2015) and Dirks (2015). These works are already rather general regarding the type of motion that is considered. In this thesis we generalize them even further by allowing a time-dependent reconstruction operator. With this step we enable the usage of the joint model in order to reconstruct data of dynamic tomography.

Tomography itself is a topic that is studied for already 100 years, since the reconstruction based on line integrals was facilitated by Radon (1917). The analysis of the Radon transform was extensively studied e.g. in Natterer (1986). However, the consideration of dynamics was so far restricted due to the massive radiation dose, which is necessary to enable proper reconstructions for every time step, cf. e.g. Niemi et al. (2013). With this work we provide an opportunity to reconstruct dynamic tomography data with a very limited amount of necessary angles.

1.4 Organization

Previously, we generally introduced and motivated the topic of this work. In this section we point out the organization of the remaining chapters. Roughly, this thesis can be divided into a theoretical part, where we derive and discuss mathematical models for tracking, and an applied part, where we use the models to solve concrete biomedical tasks. The more detailed structure of this work is described below:

We start by providing an overview of the basic principles used in this thesis. This includes a section on calculus of variations, whereas one of the most important parts states the conditions for the existence of a global minimum. Furthermore, the section treats the compactness of dual sets and presents different types of derivatives, which appear in this thesis. Continuing with a section on convex analysis, we summarize the most important details about Legendre-Fenchel Duality as well as subdifferentials. In

Section 2.3 we introduce the different function spaces, which occur in this thesis, namely Lebesgue spaces, Sobolev spaces, the space of functions with Bounded Variation, as well as Bochner spaces. In the end of the Chapter 2 we present a strategy for the numerical minimization of inverse problems.

In Chapter 3 we discuss the topic of direct tracking. In this context the expression *direct* means that we are able to directly apply a motion estimation algorithm to given data without previously needing to reconstruct the images. After introducing the topic of direct tracking, we explain the derivation of an optical flow constraint, which will be used as a data fidelity for minimization models later on. We differentiate between the standard linearized optical flow constraint and a nonlinear version, which enables to estimate motion that covers large distances. For the latter we also provide an existence and convergence analysis. We focus on different regularization strategies. On the one hand we recap the most commonly used regularizers and highlight their pros and cons. Additionally, we deduce extended regularizers, which combine positive characteristics of different standard regularizers. We also review the advantages of Bregman iterations, which are able to overcome the loss of contrast arising while using a total variation-based regularization. Afterwards, we reformulate the minimization models with respect to a primal-dual formulation, which requires the calculation of adjoint operators, convex conjugates, and resolvent operators. After a section discussing the discretization of the single terms of the models, we concentrate on the numerical realization. Therefore, we indicate the initialization and the stopping criterion that we use, as well as methods to measure the error of a calculated motion estimation result. For the concrete numerical approaches, we compare the previous data fidelities as well as the different regularizers with each other. As a last example we show some results that point out the effect of Bregman iterations.

The subsequent chapter is about indirect tracking. In contrast to Chapter 3, this chapter deals with data where the images are not directly given, but first have to be reconstructed in order to be able to apply algorithms for estimating motion. With this chapter we provide a possibility to solve both image processing tasks in a joint model. After an introduction to the topic of indirect tracking, we analyze the characteristics of a general reconstruction operator. We especially focus on the allowance of time-dependency for the operator. To get an idea for possible applications, we also give some short examples for different kinds of time-dependencies. Subsequently, we introduce a general model for indirect tracking and propose two variants for specified models. The latter include the previously studied time-dependent reconstruction as well as the optical flow constraint. For these models we also comment on the existence

of minimal solutions. Finally, we discuss the numerical realization of these models. We present an alternating minimization scheme, which realizes the reconstruction and the motion estimation tasks separately. For each of these subproblems, we provide details concerning the exact minimization algorithm.

Chapter 5 discusses the first one of two overall applications for tracking. This chapter provides a concrete task, which is supposed to be satisfied by estimating motion. This task is the analysis of intracellular motion of a certain kind of white blood cells. We introduce the biological background of this topic and point out the mathematical challenges that arise. Since the data we are dealing with requires some preprocessing before we are able to apply the motion estimation algorithms reasonably, we shortly illuminate a necessary registration step and the location of regions of interest. Finally, we apply two of the direct motion estimation models, which we derived in Chapter 3, to the real data.

The second application is the subject of Chapter 6. Here, the task requires the utilization of the indirect model discussed in Chapter 4. The application is dynamic X-ray tomography, which is introduced in the beginning of the chapter. We point out the differences between stationary and dynamic tomography, whereas we especially focus on the changes concerning the Radon transform. In general, the indirect model necessitates further a priori knowledge or initialization in order to be able to reconstruct dynamic data. For this reason, we suggest three different strategies for how to incorporate more information in a realistic manner. Those strategies contain an improved initialization for certain time steps, the usage of multiple measurements per time step, as well as the randomization of the measured directions. All of these strategies are finally applied to two different data sets. The first one is a synthetic data set, which therefore allows us to evaluate the performance of the mathematical model in terms of error measures. The second data set is a real measured data set. It serves to test the model in a realistic setting.

In Chapter 7 we summarize the content of this thesis. The chapter is divided into two parts. In the first part we concentrate on direct tracking including the results of the first application dealing with cell tracking. The second part treats the indirect tracking with the results of the second application on dynamic tomography. In each part we recap the particular analysis and draw conclusions.

The final chapter of this thesis gives a motivation for incorporating further knowledge about motion into motion estimation models. We introduce the basic principles of

mathematical modeling of motion and derive concrete examples of how to describe bidirectional movements of multiple species. The examples contain one Mean-Field model, two discrete models, and three different macroscopic models. We reformulate all of these models in a unified gradient flow formulation in order to facilitate the subsequent analysis regarding the existence of stationary solutions and the properties in regions with small densities.

2 Basics of Image Processing

In this chapter we provide the fundamental mathematical principles that serve as the background for this thesis. We start by summing up the most important results concerning the existence of minimizers and the compactness in the context of dual spaces. After defining different kinds of derivatives, we turn towards convex analysis, where we treat Legendre-Fenchel duality and subdifferential calculus. We continue by introducing several function spaces like Lebesgue spaces, Sobolev spaces, the BV space, which is of special interest in connection with image processing, and Bochner spaces. The latter are of importance when dealing with combined image reconstruction and motion estimation, which will be covered in Chapter 4.

2.1 Calculus of Variations

In this section we establish some basic definitions and results of variational calculus. After introducing operators and functionals, we summarize some fundamental results on the existence of minimizers. We continue with essential statements concerning compactness in the context of dual spaces, and end up by defining different kinds of derivatives.

The following definitions are based on Elstrodt (2006) and Rockafellar (1970).

Definition 2.1. (Measurable Space and Measure Space)
Let $\mathcal{A}$ be a σ-Algebra on a non-empty set Ω. The pair $(\Omega, \mathcal{A})$ is called *measurable space*. Furthermore, let $\mu : \mathcal{A} \to \bar{\mathbb{R}}$ be a measure. The triple $(\Omega, \mathcal{A}, \mu)$ is called *measure space*.

Definition 2.2. (Operator and Functional)
Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces with topology τ_1, τ_2 respectively. A mapping

$J : [\mathcal{X}, \tau_1) \to (\mathcal{Y}, \tau_2)$ between these spaces is called *operator*. If $\mathcal{Y}$ is a scalar field, J is called *functional*.

Definition 2.3. (Proper Functional)
Let $\mathcal{X}$ be a Banach space. A functional $J : \mathcal{X} \to \bar{\mathbb{R}}$ is called *proper*, if $J(u) \neq -\infty$ for all $u \in \mathcal{X}$ and if there exists at least one $u \in \mathcal{X}$ with $J(u) \neq \infty$.

The *effective domain* of the functional J is defined as

$$domJ := \left\{ u \in \mathcal{X} \middle| J(u) < \infty \right\}.$$

2.1.1 Existence of a Global Minimum

One of the major goals in variational calculus is to show the existence of a global minimum, which can be achieved with the help of the fundamental theorem of optimization. Before stating this theorem, we shortly define the necessary conditions, namely lower semi-continuity and coercivity. The definitions in this section are based on Ekeland and Témam (1999) and Evans (1997).

Definition 2.4. (Lower Semi-Continuity)
Let $(\mathcal{X}, \tau)$ be a measurable space and $J : (\mathcal{X}, \tau) \to \bar{\mathbb{R}}$ be a functional on $\mathcal{X}$. J is called *lower semi-continuous* in $u \in \mathcal{X}$, if

$$J(u) \leq \liminf_k J(u_k)$$

for every sequence $u_k \to u$ in the topology τ.

Definition 2.5. (Coercivity)
Let $(\mathcal{X}, \tau)$ be a measurable space and $J : (\mathcal{X}, \tau) \to \bar{\mathbb{R}}$ be a functional on $\mathcal{X}$. J is called *coercive*, if it has compact sublevel sets, i.e. there exists an $\alpha \in \mathbb{R}$ such that

$$S(\alpha) := \left\{ u \in \mathcal{X} \middle| J(u) < \alpha \right\}$$

is not empty and compact in the topology τ.

The previously described definitions of lower semi-continuity and compactness of sublevel sets lead us to the fundamental theorem of optimization. See e.g. Aubert and Kornprobst (2006) for further details of the theorem.

Theorem 2.1. (Fundamental Theorem of Optimization)
Let $(\mathcal{X}, \tau)$ be a metric space and $J : (\mathcal{X}, \tau) \to \bar{\mathbb{R}}$ be a functional on $\mathcal{X}$. Furthermore let

J be lower semi-continuous (Definition 2.4) and coercive (Definition 2.5).
Then there exists a global minimum $\bar{u} \in \mathcal{X}$, i.e.

$$J(\bar{u}) = \inf_{u \in \mathcal{X}}(u).$$

Proof. Let $(u_k)_{k\in\mathbb{N}}$ be a minimizing sequence, i.e. $J(u_k) \to \inf_{u\in\mathcal{X}} J(u)$. Let $k_0 \in \mathbb{N}$ be sufficiently large such that $u_{k_0} \in S(\alpha)$ (Definition 2.5). Then $(u_k)_{k\geq k_0}$ is contained in a compact set and has a convergent subsequence $(\hat{u}_k)_{k\in\mathbb{N}}$ with $J(\hat{u}_k) \to \inf_{u\in\mathcal{X}} J(\hat{u})$. Due to the lower semi-continuity we achieve

$$\inf_{u\in\mathcal{X}} J(u) \leq J(\hat{u}) \underbrace{\leq}_{\text{(Definition 2.4)}} \lim_{k\to\infty} = J(\hat{u}_k) = \inf_{u\in\mathcal{X}} J(u).$$

Consequently, $\hat{u}$ is a global minimum of J.

□

2.1.2 Compactness of Dual Sets

Compact sets are of great importance for many areas of functional analysis. However, since they are not always given it is often useful to switch over to the weak*-topology, for which the compactness can be achieved with the Theorem of Banach-Alaoglu. In order to declare this theorem, we previously state some required definitions, which are based on Ekeland and Témam (1999) and Dacorogna (2004).

Definition 2.6. (Dual Space)
The *dual space* $\mathcal{X}^*$ of a metric linear space $\mathcal{X}$ is defined as the space of all continuous linear functionals $J : \mathcal{X} \to \mathbb{R}$.

$\mathcal{X}^*$ is a normed space. The norm in $\mathcal{X}^*$ is given by

$$\|J\|_{\mathcal{X}^*} = \sup_{u\in\mathcal{X}\setminus\{0\}} \frac{|J(u)|}{\|u\|_{\mathcal{X}}}.$$

Definition 2.7. (Reflexive Space)
Let $\mathcal{X}$ be a metric linear space with dual space $\mathcal{X}^*$. The bidual space $\mathcal{X}^{**}$ of $\mathcal{X}$ is defined as the dual space of $\mathcal{X}^*$. The space $\mathcal{X}$ is called *reflexive*, if the evaluation map $J : \mathcal{X} \to \mathcal{X}^{**}$ is an isometric isomorphism.

Definition 2.8. (Weak and Weak*-Topology)
Let $\mathcal{X}$ be a Banach space with dual space $\mathcal{X}^*$. The *weak topology* on $\mathcal{X}$ is defined by

$$u_k \rightharpoonup u :\Leftrightarrow \langle v, u_k\rangle \to \langle v, u\rangle \quad \forall v \in \mathcal{X}^*.$$

The *weak* topology* on $\mathcal{X}^*$ is defined by

$$v_k \rightharpoonup^* v :\Leftrightarrow \langle v_k, u\rangle \to \langle v, u\rangle \quad \forall u \in \mathcal{X}.$$

Remark 2.1. The weak and the weak* topology coincide in reflexive Banach spaces, i.e. if $\mathcal{X}^{**} = \mathcal{X}$.

Lemma 2.1. (Weak* Lower Semi-Continuity in Banach Spaces)
Let $\mathcal{X}$ be a Banach space with dual space $\mathcal{X}^*$. Then the norm $\|\cdot\|_{\mathcal{X}}$ of the Banach space $\mathcal{X}$ is weak* lower semi-continuous.

Proof. Let $v_k \in \mathcal{X}^*$ be a sequence that converges towards $v \in \mathcal{X}$ in the weak* topology, i.e.

$$\langle v_k, u\rangle \to \langle v, u\rangle \quad \forall u \in \mathcal{X}.$$

For the first scalar product it holds

$$|\langle v_k, u\rangle| \leq \|v_k\|_{\mathcal{X}^*}\|u\|_{\mathcal{X}}.$$

Taking the limit $k \to \infty$ we obtain

$$|\langle v, u\rangle| \leq \liminf_{k\to\infty} \|v_k\|_{\mathcal{X}^*}\|u\|_{\mathcal{X}}.$$

Since this inequality holds for every $u \in \mathcal{X}$, it follows

$$\|v\|_{\mathcal{X}} = \sup_{\|u\|\leq 1} \langle v, u\rangle \leq \liminf_{k\to\infty} \|v_k\|_{\mathcal{X}^*}.$$

□

Theorem 2.2. (Theorem of Banach-Alaoglu)
Let $\mathcal{X}$ be a Banach space with dual space $\mathcal{X}^*$. For $C > 0$ the set

$$\left\{v \in \mathcal{X}^* \middle| \|v\|_{\mathcal{X}^*} \leq C\right\}$$

is compact in the weak* topology.

A proof can e.g. be found in Rudin (1973), p. 66-68, Chapter 3, Theorem 3.15.

2.1.3 Derivatives

In the course of this thesis we will require different types of derivatives for functionals. In the following we specify the directional derivative, the Gâteaux derivative, as well as the Fréchet derivative. These definitions are mainly based on Evans (1997).

Definition 2.9. (Directional Derivative)
Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces and $J : \mathcal{X} \to \mathcal{Y}$ be a proper functional. If the limit exists, the *directional derivative* at the point $u \in \mathcal{X}$ in the direction v is defined as

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

If the directional derivative exists for all $u \in \mathcal{X}$, J is called *directionally differentiable*.

Definition 2.10. (Gâteaux Derivative)
Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces and $J : \mathcal{X} \to \mathcal{Y}$ be a proper functional. The *Gâteaux derivative* in $u \in \mathcal{X}$ is defined as the set of directional derivatives, i.e.

$$dJ(u) := \{d_v J(u) \big| d_v J(u) < \infty \text{ for } v \in \mathcal{X}\}.$$

If $dJ(u) \neq \{\emptyset\}$, J is called *Gâteaux differentiable* in u.

Definition 2.11. (Fréchet Derivative)
Let $\mathcal{X}$ and $\mathcal{Y}$ be two Banach spaces and $J : \mathcal{X} \to \mathcal{Y}$ be a proper and Gâteaux differentiable functional in $u \in \mathcal{X}$. The *Fréchet derivative* in u is defined as a continuous linear functional $F : \mathcal{X} \to \mathcal{Y}$ with

$$Fv = d_v J(u) \quad \forall v \in \mathcal{X},$$

and

$$\lim_{\|v\|_{\mathcal{X}} \to 0} \frac{|J(u+v) - J(u) - Fv|}{\|v\|_{\mathcal{X}}} = 0 \quad \forall v \in \mathcal{X}.$$

If such an F exists, J is called *Fréchet differentiable* at u.

2.2 Convex Analysis

In this section we illustrate some crucial results concerning convex analysis. After defining convex sets and convex functionals, we treat Legendre-Fenchel duality as well as the extended definition of differentiability. For a detailed study on this topic we

refer the interested reader to Rockafellar (1970), which is also the basis for the definitions in this section.

Definition 2.12. (Convex Subset)
Let $\mathcal{X}$ be a Banach space. A subset $X \subset \mathcal{X}$ is called *convex*, if

$$\lambda u + (1 - \lambda)v \in X$$

for all $u, v \in X$ and all $\lambda \in [0, 1]$.

Definition 2.13. (Convex Functional)
Let $\mathcal{X}$ be a Banach space with convex subset $X \subset \mathcal{X}$. A functional $J : X \to \bar{\mathbb{R}}$ is called *convex*, if

$$J\left(\lambda u + (1 - \lambda u)v\right) \leq \lambda J(u) + (1 - \lambda)J(v)$$

for all $u, v \in X$ and all $\lambda \in [0, 1]$.
J is called *strictly convex*, if the above inequality is strict for all $\lambda \in (0, 1)$ and $u \neq v$.

2.2.1 Legendre-Fenchel Duality

The minimization of functionals based on their primal formulations can be quite complex in some cases. Often, it is more efficient to study the dual formulation instead. For this reason, we give a short overview about some essential results concerning duality.

Definition 2.14. (Convex Conjugate)
Let $\mathcal{X}$ be a Banach space with dual space $\mathcal{X}^*$. The *convex conjugate* $J^* : \mathcal{X}^* \to \bar{\mathbb{R}}$ of a functional $J : \mathcal{X} \to \bar{\mathbb{R}}$ is defined by

$$J^*(p) := \sup_{u \in \mathcal{X}} \{\langle p, u \rangle_{\mathcal{X}} - J(u)\}$$

for $p \in \mathcal{X}^*$.
The *biconjugate* $J^{**} : \mathcal{X} \to \bar{\mathbb{R}}$ of J is defined by

$$J^{**}(u) := \sup_{p \in \mathcal{X}^*} \{\langle u, p \rangle_{\mathcal{X}^*} - J^*(p)\}$$

for $u \in \mathcal{X}$.

The convex conjugate is also known as *Legendre-Fenchel transform*, which was named after Adrien-Marie Legendre and Werner Fenchel.

Lemma 2.2. (Properties of the Convex Conjugate)
Let $\mathcal{X}$ be a Banach space with dual space $\mathcal{X}^*$ and $J : \mathcal{X} \to \mathbb{R}$ be a proper, convex and lower semi-continuous functional. For $p \in \mathcal{X}^*$ the following properties hold:

1. $(J(p+a))^* = J^*(p) - \langle a, p\rangle, \quad a \in \mathbb{R}.$
2. $(J(p)+a)^* = J^*(p) - a, \quad a \in \mathbb{R},$
3. $(\alpha J(p))^* = \alpha J^*\left(\frac{p}{\alpha}\right), \quad \alpha > 0.$

Proof. 1. Inserting the requested functional into the definition of the convex conjugate and additionally using the equivalent dual formulation of the occurring scalar product of a and u yields

$$\begin{aligned}(J(p+a))^* &= \sup_{u\in\mathcal{X}} \{\langle p+a, u\rangle - J(u)\} \\ &= \sup_{u\in\mathcal{X}} \{\langle p, u\rangle - J(u) + \langle a, u\rangle\} \\ &= \sup_{u\in\mathcal{X}} \{\langle p, u\rangle - J(u) - \langle a, p\rangle\} \\ &= J^*(p) - \langle a, p\rangle.\end{aligned}$$

2. We simply insert $(J(p)+a)^*$ into the definition of the convex conjugate directly and obtain

$$\begin{aligned}(J(p)+a)^* &= \sup_{u\in\mathcal{X}} \{\langle p, u\rangle - (J(u)+\alpha)\} \\ &= \sup_{u\in\mathcal{X}} \{\langle p, u\rangle - J(u) - \alpha\} \\ &= J^*(p) - a.\end{aligned}$$

3. Again, we simply insert $(\alpha J(p))^*$ into the definition of the convex conjugate to achieve

$$\begin{aligned}(\alpha J(p))^* &= \sup_{u\in\mathcal{X}} \{\langle p, u\rangle - \alpha J(u)\} \\ &= \alpha \sup_{u\in\mathcal{X}} \left\{\left\langle \frac{p}{\alpha}, u\right\rangle - J(u)\right\} \\ &= \alpha J^*\left(\frac{p}{\alpha}\right).\end{aligned}$$

□

The following theorem gives a necessary and sufficient criterion for the functional J to be equal to its biconjugate J^{**}. This condition can be used to show that the solution of a primal problem is equal to the solution of its dual formulation.

Theorem 2.3. (Fenchel-Moreau-Rockafellar Theorem)
Let $\mathcal{X}$ be a Banach space. For $J : \mathcal{X} \to \bar{\mathbb{R}}$ it holds

1. $J^{**} \leq J$,
2. $J^{**} = J$ iff J is convex and lower semi-continuous.

A proof can e.g. be found in Rockafellar (1970).

The biconjugate is thus the *closed convex hull*, i.e. the largest lower semi-continuous convex functional smaller than J.

The following theorem is one of the fundamental results in convex analysis. It allows to reformulate a primal variational problem into an equivalent dual problem. In numerous cases, the minimization of the dual problem is easier to handle. This leads to a great importance of the theorem for many numerical methods for variational problems.

Theorem 2.4. (Fenchel's Duality Theorem)
Let $\mathcal{X}$ and $\mathcal{Y}$ and $J_1 : \mathcal{X} \to \bar{\mathbb{R}}$ and $J_2 : \mathcal{Y} \to \bar{\mathbb{R}}$ be proper, lower semi-continuous and convex functionals with $domJ_1 \cap domJ_2 \neq \emptyset$. Furthermore, let $K : \mathcal{X} \to \mathcal{Y}$ be a bounded linear operator. Then it holds

$$\inf_{u \in \mathcal{X}} \{J_1(Ku) + J_2(u)\} = \sup_{p \in \mathcal{X}^*} \{J_1^*(p) + J_2^*(Kp)\}.$$

A proof ca e.g. be found in Ekeland and Témam (1999), Chapter 3, Sections 1-4. Duality will be of great importance in the following parts of this thesis. To obtain a more detailed understanding of this issue, the reader can find some descriptive examples in the context of image processing in Dirks (2015).

2.2.2 Subdifferentials

For the subsequent analysis of variational problems we want to be able to consider more general functions, which are not necessarily Fréchet differentiable in every point. For this sake, we extend the definition of the Fréchet derivative and introduce the subdifferential of a convex function.

Definition 2.15. (Subdifferential)
Let $\mathcal{X}$ be a Banach space with dual space $\mathcal{X}^*$ and $J : \mathcal{X} \to \bar{\mathbb{R}}$ be a proper and convex functional. J is called *subdifferentiable* in $u \in \mathcal{X}$, if there exists a $p \in \mathcal{X}^*$ such that

$$J(v) - J(u) - \{p, v - u\} \geq 0 \quad \forall v \in \mathcal{X}.$$

p is called the *subgradient* of J in u. The *subdifferential* of J in u is defined as the set of all subgradients in u, i.e.

$$\partial J(u) \coloneqq \left\{ p \in \mathcal{X}^* \middle| J(u) - J(v) - \{p, v-u\} \geq 0 \, \forall v \in \mathcal{X} \right\}.$$

2.3 Function Spaces

In this section we introduce the function spaces that will be most important for the subsequent analysis in this thesis. We start this section with Lebesgue and Sobolev spaces. Since total variation will be a fundamental part of this thesis, we proceed with the space of functions with bounded variations. Finally, we specify Bochner spaces, which will be used in the context of indirect tracking in Chapter 4.

2.3.1 Lebesgue Spaces

In the following, we shortly summarize the definition of the Lebesgue spaces as well as of their norms and some basic properties. Further details about this topic can e.g. be found in Rudin (1987). The following definition is based on Dacorogna (2004).

Definition 2.16. (Lebesgue Space)
The *Lebesgue space* $L^p(\Omega)$ for $1 \leq p < \infty$ is defined as the set of all measurable functions $u : \Omega \to \mathbb{C}$ on the domain Ω for which holds

$$\int_\Omega |u(x)|^p dx < \infty.$$

For $p = \infty$ the Lebesgue space $L^\infty(\Omega)$ is defined as the set of all measurable functions $u : \Omega \to \mathbb{C}$ on the domain Ω for which holds

$$\operatorname{ess\,sup} |u| < \infty,$$

where the essential supremum ess sup of u is the supremum except for a set $\mathcal{N}$ of measure zero.

For $1 \leq p < \infty$ the norm in $L^p(\Omega)$ is given by

$$\|u\|_{L^p(\Omega)} \coloneqq \left(\int_\Omega |u|^p \right)^{\frac{1}{p}}.$$

For $p = \infty$ the norm in $L^\infty\Omega$ is given by

$$\|u\|_{L^\infty(\Omega)} \coloneqq \operatorname{ess\,sup} |u| = \inf_{\mathcal{N}} \sup_{x \in \Omega \setminus \mathcal{N}} |u(x)|.$$

In the remainder of the thesis we will shorten the notation of the above norms by writing $\|u\|_p$ instead of $\|u\|_{L^p(\Omega)}$ for $p \geq 1$.

Remark 2.2. For $p \geq 1$ the Lebesgue space $L^p(\Omega)$ is a Banach space with norm $\|u\|_p$.

Remark 2.3. In the special case of $p = 2$ the Lebesgue space $L^2(\Omega)$ is a Hilbert space with the scalar product

$$\langle u, v \rangle_{L^2(\Omega)} := \int_\Omega \langle u(x), v(x) \rangle \, dx.$$

Remark 2.4. For $1 < p < \infty$ the dual space of $L^p(\Omega)$ is again a Lebesgue space, i.e. $(L^p(\Omega))^* = L^q(\Omega)$, where $\frac{1}{p} + \frac{1}{q} = 1$. Consequently, these spaces are reflexive. For $p = 2$ the dual space of $L^2(\Omega)$ is again $L^2(\Omega)$. If the metric space $(\Omega, \mathcal{A}, \mu)$ is σ-finite, the dual space of $L^1(\Omega)$ is given by $L^\infty(\Omega)$. Nevertheless, $L^1(\Omega)$ and $L^\infty(\Omega)$ are not reflexive.

2.3.2 Sobolev Spaces

The previously introduced Lebesgue spaces include a wide range of function types, amongst others functions with high oscillations. The latter usually indicate noise and are thus not desirable. To avoid such noisy functions, it is possible to only study those functions whose derivatives are Lebesgue integrable. A space that only contains such functions is the Sobolev space. The definition of the Sobolev space is based on Evans (1997).

Definition 2.17. (Sobolev Space)
Let u be a measurable function on a domain Ω and $1 \leq k, p < \infty$. The *Sobolev space* $W^{k,p}(\Omega)$ is defined as

$$W^{k,p}(\Omega) = \left\{ u \in L^p(\Omega) \middle| D^\alpha u \in L^p(\Omega) \forall |\alpha| \leq k \right\}.$$

The norm in $W^{k,p}(\Omega)$is given by

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

Again, to shorten the notation of the norm we write $\|u\|_{k,p}$ instead of $\|u\|_{W^{k,p}(\Omega)}$ in the remainder of the thesis.

Remark 2.5. The Sobolev space $W^{k,p}(\Omega)$ is a Banach space with norm $\|u\|_{k,p}$.

Remark 2.6. For $p = 2$, $W^{k,2}(\Omega)$ is a Hilbert space with the scalar product

$$\langle u, v \rangle_{W^{k,p}(\Omega)} := \sum_{|\alpha| \leq k} \langle D^\alpha u, D^\alpha v \rangle_{L^2(\Omega)} .$$

For $k \in \mathbb{N}$, the Hilbert space $W^{k,2}(\Omega)$ is usually denoted by

$$H^k(\Omega) := W^{k,2}(\Omega).$$

2.3.3 Space of Functions with Bounded Variation

In this section we introduce the space of functions with bounded variations. Thanks to some characteristics being exceedingly useful in the context of image processing, this space will be of special interest in the remaining parts of this thesis. Often, images have typical characteristics that should be considered while choosing a proper function space for the analysis. One of the typical requirements for the image space is that it contains discontinuities. This is important, since discontinuities depict sharp edges, which on the other hand define e.g. an object in an image. A second typical requirement to the space is that it avoids noise, which is usually undesirable. A function space that satisfies the latter is e.g. the Sobolev space $W^{1,1}(\Omega)$. This space includes smooth images and therefore excludes the appearance of noise. Unfortunately, it does not fulfill the first requirement to contain discontinuities. This can be satisfied by using the larger Lebesgue space $L^1(\Omega)$. This however includes noise. Consequently, a desirable function space is between the Sobolev and the Lebesgue spaces and this is exactly where the space of function with bounded variations is located. For further details about this space we refer to Ambrosio et al. (2000) and Burger and Osher (2013).

In the remainder of this section we assume $\Omega \subset \mathbb{R}^N$ to be open and sufficiently regular. To clarify the idea of the space, we start with the definition of the total variation.

Definition 2.18. (Total Variation)
Let $u \in L^1(\Omega)$ be a function on a domain Ω. The *total variation* (TV) of u is defined as

$$TV(u) := \sup_{\substack{\phi \in C_0^\infty, \\ \|\phi\|_\infty \leq 1}} \int_\Omega u \nabla \cdot \phi \, dx.$$

Definition 2.19. (BV Space)
Let Ω be a domain. The *space of functions with bounded total variation* (BV) is defined

as

$$BV(\Omega) \coloneqq \left\{ u \in L^1(\Omega) \middle| TV(u) < \infty \right\}.$$

Remark 2.7. The total variation $TV(u) = |u|_{BV}$ is a semi-norm on $BV(\Omega)$.

Remark 2.8. The BV space is a Banach space with the norm

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

Remark 2.9. If the function u is sufficiently smooth, i.e. $u \in W^{1,1}(\Omega)$, a primal formulation of $TV(u)$ is given by

$$TV(u) = \int_\Omega |\nabla u| \, dx.$$

Since $W^{1,1}(\Omega)$ is a subspace of $BV(\Omega)$, in informal settings this formulation is often used for the general function space $BV(\Omega)$. In fact, the inclusion is strict and the formulation is not proper for every function in $BV(\Omega)$. An easy example for a misfit is the Heaviside

$$H(u) = \begin{cases} 1, & \text{if } u \geq 0 \\ 0, & \text{else} \end{cases},$$

which is contained in $BV(\Omega)$, but not in $W^{1,1}(\Omega)$.

Lemma 2.3. (Embedding of $BV(\Omega)$)
For $\Omega \subset \mathbb{R}^N$ and $\frac{p}{p-1} \geq N$ it holds

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

For $\frac{p}{p-1} > N$ the embedding is compact.

A proof can e.g. be found in Giusti (1984).

Lemma 2.4. (Convexity of TV)
The total variation $TV(u)$ for $u \in L^1(\Omega)$ is convex.

Proof. Choose arbitrary $u, v \in L^1(\Omega)$ and $\lambda \in [0,1]$. Then it holds

$$\begin{aligned} TV(\lambda u + (1-\lambda)v) &= \sup_{\substack{\phi \in C_0^\infty, \\ \|\phi\|_\infty \leq 1}} \int_\Omega (\lambda u + (1-\lambda)v) \nabla \cdot \phi \, dx \\ &\leq \lambda \sup_{\substack{\phi \in C_0^\infty, \\ \|\phi\|_\infty \leq 1}} \int_\Omega u \nabla \cdot \phi \, dx + (1-\lambda) \sup_{\substack{\phi \in C_0^\infty, \\ \|\phi\|_\infty \leq 1}} \int_\Omega v \nabla \cdot \phi \, dx \\ &= \lambda TV(u) + (1-\lambda) TV(v). \end{aligned}$$

□

Remark 2.10. $|\cdot|_{BV}$ is lower semi-continuous with respect to the weak topology of $L^1(\Omega)$. Furthermore, all uniformly bounded sequences in $BV(\Omega)$ are relatively compact in $L^1(\Omega)$. Both arguments can be seen e.g. in Acar and Vogel (1994).

2.3.4 Bochner Spaces

The usual Lebesgue spaces are only defined for functions with values in $\mathbb{C}$. The Bochner spaces represent a generalization for functions with values in an arbitrary Banach space, which not necessarily needs to be the Banach space $\mathbb{C}$. The following definitions are based on Showalter (2013).

Definition 2.20. (Weak Measurability)
Let $\mathcal{X}$ be a Banach space. A function $J : [0,T] \to \mathcal{X}$ is called *weakly measurable*, if the map $t \mapsto (J(u), u)_{\mathcal{X}}$ is measurable on $[0,T]$ for every $u \in \mathcal{X}$.

Definition 2.21. (Strong Measurability)
Let $\mathcal{X}$ be a Banach space. A function $J : [0,T] \to \mathcal{X}$ is called *strongly measurable*, if it holds

$$J(t) = \lim_{n \to \infty} J_n(t) \quad \text{for almost every } t,$$

where the functions $(J_n)_{n \in \mathbb{N}}$ each have a countable range and the inverse $J^{-1}\{u\}$ of J is measurable for each $u \in \mathcal{X}$.

Strong measurable functions are also called *Bochner measurable*.

Remark 2.11. Strong measurability implies weak measurability. Due to the Orlicz–Pettis Theorem (Orlicz (1929), Pettis (1938)), both definitions are equivalent in case that the Banach space $\mathcal{X}$ is separable.

Definition 2.22. (Bochner Space)
Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|_{\mathcal{X}}$ and $1 \leq p < \infty$. The *Bochner space* $L^p([0,T];\mathcal{X})$ consists of equivalence classes of strongly measurable functions $u : [0,T] \to \mathcal{X}$ with norm

$$\|u\|_{L^p([0,T];\mathcal{X})} := \left(\int_0^T \|u(t)\|_{\mathcal{X}^p} \, dt \right)^{\frac{1}{p}} < \infty$$

and

$$\|u\|_{L^\infty([0,T];\mathcal{X})} := \operatorname*{ess\,sup}_{0 \leq t \leq T} \|u(t)\| < \infty.$$

2.4 Numerical Optimization

In this section we introduce a framework for variational minimization of convex problems. Within a short period of time, similar frameworks were proposed independently by several authors like Esser et al. (2010), Chambolle and Pock (2011) and Zhang et al. (2011). They all combine the idea to take advantage of the fact that often parts of an energy functional are easier so solve by using an equivalent dual formulation. This is why the frameworks are usually known as *primal-dual* methods. The most popular framework in this category is the one proposed by Chambolle and Pock. This is also the one that we will introduce in this section. The algorithm is written in a general manner and can be adapted to a wide range of variational problems. All of the algorithms that will be processed in this thesis are based on this framework.

As a requirement for the framework we assume to have two finite-dimensional real vector spaces $\mathcal{X}$ and $\mathcal{Y}$ equipped with an inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{\frac{1}{2}}$. Moreover, we consider a continuous linear operator $K : \mathcal{X} \to \mathcal{Y}$ with induced norm

$$\|K\| = \max \left\{ \|Ku\| \middle| u \in \mathcal{X} \text{ with } \|u\| \leq 1 \right\}.$$

In general, the problems we want to analyze in this thesis can be written as

$$\min_{u \in \mathcal{X}} F(Ku) + G(u), \tag{2.1}$$

where $F, G : \mathcal{X} \to \bar{\mathbb{R}}$ are proper, convex, lower semi-continuous functions. This is a primal formulation of a general energy functional. This problem is equivalent to the

corresponding dual formulation

$$\max_{p\in\mathcal{Y}} - \left(G^*(-K^*p) + F^*(p)\right),$$

where F^* is the convex conjugate of F, as well as to the saddle point problem

$$\min_{u\in\mathcal{X}} \max_{p\in\mathcal{Y}} G(u) + \langle u, K^*p\rangle - F^*(p). \tag{2.2}$$

The latter is the primal-dual problem and also the basis of the framework that we introduce. We assume that the above problems have at least one solution $(\hat{u}, \hat{p}) \in \mathcal{X}\times\mathcal{Y}$ that satisfies

$$K\hat{u} \in \partial F^*(\hat{p}) \qquad \textit{and} \qquad -(K^*\hat{p}) \in \partial G(\hat{u}),$$

where ∂F^* and ∂G are the subgradients of the convex functions F^* and G. We define the resolvent operators for such functions as follows:

Definition 2.23. (Resolvent Operator)
Let $\mathcal{X}$ be a Banach space and $F : \mathcal{X} \to \bar{\mathbb{R}}$ be a proper functional. For $\tau > 0$ the *resolvent operator* is defined as

$$u = (I + \tau\partial F)^{-1}(\tilde{u}) \coloneqq \arg\min_u \left\{ \frac{\|u-\tilde{u}\|^2}{2\tau} + F(u) \right\}.$$

To calculate the resolvent operator for a dual functional, we need the following theorem:

Theorem 2.5. (Moreau's Identity)
Let $\mathcal{X}$ be a Banach space and $F : \mathcal{X} \to \bar{\mathbb{R}}$ be a proper functional with corresponding dual functional F^*. The resolvent operators for the primal and the dual functional are linked in the following way:

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

A proof can e.g. be found in Showalter (2013). Thus, the resolvent operator for the dual functional F^* can be written as

$$\left(I + \frac{1}{\tau}\partial F^*\right)^{-1}(u) \coloneqq u - \frac{1}{\tau}(I + \tau\partial F)^{-1}(\tau u).$$

With the above assumptions, the primal-dual problem (2.2) can be solved by alternating fixed-point iterations for u and p. With an additional relaxation step for u, we obtain the following minimizing algorithm:

Definition 2.24. (Chambolle-Pock Algorithm)
Choose $\tau, \sigma > 0$, $\theta \in [0,1]$, $(u^0, p^0) \in \mathcal{X} \times \mathcal{Y}$ and set $\bar{u}^0 = u^0$. For $n \geq 0$ the updates of u^n, p^n and $\bar{u^n}$ are defined as follows:

$$\begin{aligned} p^{n+1} &= (I + \sigma \partial F^*)^{-1}(p^n + \sigma K \bar{u}^n), \\ u^{n+1} &= (I + \tau \partial G)^{-1}(u^n - \tau K^* p^{n+1}), \\ \bar{u}^{n+1} &= u^{n+1} + \theta(u^{n+1} - u^n). \end{aligned}$$

Remark 2.12. For $\tau\sigma\|K\|^2 < 1$ and $(u^n, \bar{u}^n, p^n)$ as in Definition 2.24 there exists a saddle point (u^*, p^*) such that $u^n \to u^*$ and $p^n \to p^*$. A proof can be found in Chambolle and Pock (2011).

Remark 2.13. For $K = I$ the Chambolle-Pock Algorithm is amongst others equivalent to the Douglas-Rachford Splitting Algorithm (Douglas and Rachford (1956)), as well as to the alternating direction method of multipliers (ADMM) (see e.g. Boyd et al. (2011)). For general K Definition 2.24 is equivalent to a preconditioned version of the ADMM. A nice overview of several splitting methods and their connections can be found in Brune (2010).

3 Direct Tracking

In this chapter we introduce methods to track objects over multiple time steps. The focus is on estimating motion based on optical flow. An application for the methods introduced in this section can be found in Chapter 5.

We start by specifying the idea behind computational motion, at which we also explain some typical problems that arise during the estimation. In Section 3.2 we present a variational approach for motion estimation that contains the optical flow constraint as a data fidelity. Afterwards, we complete the model by adding different kinds of regularizers in Section 3.3. In Section 3.3.1 we use standard regularization terms like TV or the L^2 norm. Since it turns out that those standard regularizers only fit to specific motion estimation tasks, we generalize the regularizers by adding an extra term in Section 3.3.2. This leads to an extended version of the TV regularizer. Subsequently, in Section 3.4 we also introduce Bregman iterations as a possibility to solve the problem of loss of contrast that occurs in models that contain TV-based regularizers. In Section 3.5 we reformulate the introduced models as saddle point problems, since these are required for the numerical implementation that we describe in Section 3.7. Here, we describe the numerical algorithm, the type of discretization, as well as the choice of parameters. Finally, we evaluate the different models based on synthetic examples.

3.1 Introduction

The tracking of objects or certain structures is dependent on apparent motion. The human brain is able to calculate motion based on a global view of a scene. In this chapter we estimate the apparent motion between images taken at consecutive time steps. The problem with motion estimation in image processing is the so-called aperture

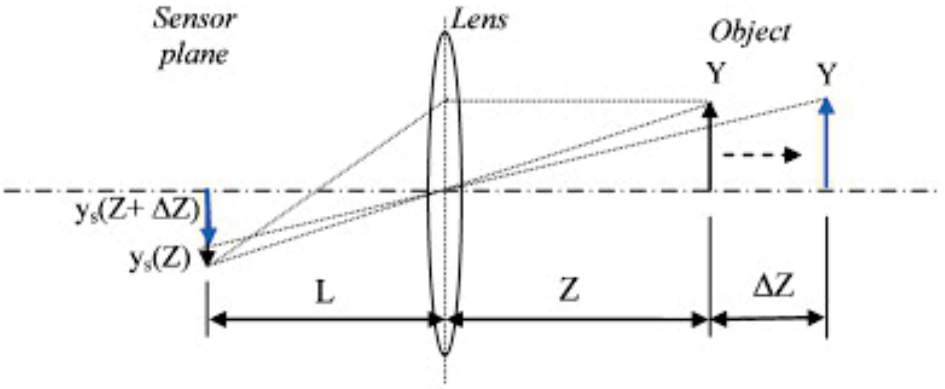

Figure 3.1: Difference between projected and real motion, from SUTTON et al. (2008).

problem, i.e. the missing feasibility of getting a global view of a scene. The main reason for this problem is that an image is usually two-dimensional. The movement on the other hand takes place in three dimensions. In images it is not possible to uniquely identify motion that takes place in the same direction as the angle of view of the image. E.g., it is not possible to decide whether an object grows but stays at the exact same position, or if it approaches but keeps its size. Figure 3.1 visualizes the difference of the real three-dimensional motion and the two-dimensional projection onto a camera lens. The human eyes are partly able to overcome this problem through binocular vision, i.e. by seeing each scene through both eyes separately. The brain combines both perspectives to obtain a three-dimensional perception. However, if the view is limited in x- and y-direction, the aperture problem occurs even if the motion is restricted to two dimensions. It is possible that different kinds of movement with different orientations result in the exact same images. See Figure 3.2 for an illustration of this issue. A similar effect occurs if a person is looking out of a window of a stationary train and everything this person sees is an oncoming train. From his perspective it is not possible to say whether his train is moving backwards, or if the other train is moving forwards. Thus, the apparent movement is not necessarily identical to the real motion. This non-uniqueness is a reason why motion estimation is an ill-posed problem, which makes it challenging to solve.

Image processing is always performed on digital images that consist of pixels with different intensities. These intensities are also the only basis for estimating motion. The goal is to connect pixels in consecutive time steps based on their intensities. Typical image processing problems like noise, low contrast etc. impede this task. They usually occur independently in every image, which makes it even more difficult to connect the correct pixels in different images. However, estimating motion additionally contains some specific challenges. One problem is that the illumination of the displayed scene can

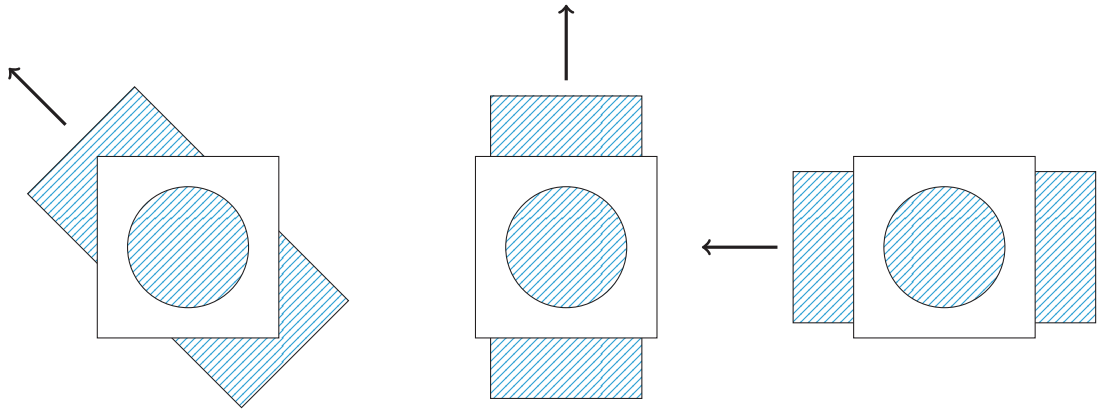

Figure 3.2: Aperture problem: the motion within the circle is not distinguishable.

change over time. Two pixels in consecutive images can have totally different intensities, even if they depict the exact same point, since there is e.g. a shadow over the scene at one of the time steps. Another special difficulty is the temporal resolution. In the majority of cases it is useful to prefer a connection of pixels that are close together, since everything else would destroy the structure of the image. However, this only makes sense if the movement is relatively small. The distance of movement is directly connected to the time that passes between the images and the speed of motion. Thus, it is always desirable to have a high temporal resolution to receive motion as small as possible. However, especially in a medical context it can take several minutes to get a single image.

For the remaining thesis we denote the domain of an image by Ω. Since an image has typically two or three dimensions, we assume $\Omega \in \mathbb{R}^d$ for $d = 2, 3$. The time interval we consider is referred to as $[0, T]$. The variable $u(x, t)$ describes the intensity of an image at the spatial position $x \in \Omega$ at time $t \in [0, T]$, i.e. $u : \Omega \times [0, T] \to \mathbb{R}^+$. We denote the flow field that we aim to calculate by $\boldsymbol{v}$. This variable separately assigns the velocity to every pixel in every spatial dimension. Thus, it is a vector with d components, i.e. $\boldsymbol{v} : \mathbb{R}^d \to \mathbb{R}^d$.

3.2 The Optical Flow Constraint

There are different strategies to estimate motion between consecutive images. Every single method has its own pros and cons, which depend on the kind of motion that we aim to detect. In this thesis we focus on one of the most popular strategies for motion estimation, which is called optical flow. The basic idea of optical flow is the brightness constancy assumption. This means that the pixels of every object have the same intensity independent of the particular time step. Thus, the time derivative of the image intensity equals zero. By connecting identical pixels this way, it is possible to

find movement between two images. Certainly, this brightness constancy assumption does not hold for every kind of data. E.g., if the illumination changes over time, the approach does not permit an accurate calculation of the motion. However, we stick to the proposed optical flow approach throughout this thesis, since it works well for a wide range of applications. See e.g. BERGEN et al. (1992) for an overview of different motion estimation models and their particular characteristics.

In the subsequent section we describe different strategies of how to calculate optical flow. We start with the most commonly used linear approach. This however has heavy restrictions concerning the distance of the movement. We continue by describing strategies of how to avoid this problem but still use the linear approximation. Afterwards, we also introduce a nonlinear approach, which is also able to estimate movement with larger distances in Section 3.2.2.

3.2.1 Linearized Approach

The brightness constancy assumption can be explained as a constant intensity of the single images along the trajectory $x(t)$, i.e. the total derivative of $u(x,t)$ with respect to t shall be zero:

$$0 = \frac{du}{dt}.$$

Using the chain rule we obtain:

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \sum_{i=1}^{d} \frac{\partial u}{\partial x_i}\frac{dx_i}{dt}.$$

The temporal derivative $\frac{dx_i}{dt}$ of the trajectory $x(t)$ is identical to the velocity field $\boldsymbol{v}(x)$ that we are looking for. Additionally, the spatial derivatives $\sum_{i=1}^{d} \frac{\partial u(x_i)}{\partial x_i}$ are equivalent to the gradient operator $\nabla u(x_i)$. Furthermore, replacing $\frac{\partial u}{\partial t}$ by the shorter notation u_t and considering that the velocity field $\boldsymbol{v}$ is given by the temporal derivative of the trajectory, i.e. $\boldsymbol{v}(x) = \frac{dx}{dt}$, we receive the following *linearized optical flow constraint*:

$$0 = u_t(x,t) + \nabla u(x,t) \cdot \boldsymbol{v}(x). \tag{3.1}$$

The linearized optical flow constraint can also be regarded as an inverse problem $A\boldsymbol{v} = g$ with $A : L^p(\Omega)^2 \to L^p(\Omega)$ for $p \geq 1$, by defining

$$A\boldsymbol{v} = \nabla u \cdot \boldsymbol{v} \qquad \text{and} \qquad g = -u_t.$$

See Aubert and Kornprobst (2006) or Becker et al. (2015) to find a more detailed derivation of the optical flow constraint.

In (3.1) it is shown that for every pixel or voxel x we need to find solutions for every single component of $\boldsymbol{v}$. In one dimension this can work quite well, since we have just as many equations as unknowns. However, a regular image usually has two or even three dimensions, which results in a highly under-determined system of equations. This is a consequence of the previously described aperture problem. It prevents us from finding a unique solution by only regarding the optical flow constraint. Thus, it is necessary to incorporate a priori information, which we realize by using a variational approach of the form

$$\min_{\boldsymbol{v}} \mathcal{D}(u, \boldsymbol{v}) + \alpha \mathcal{R}(\boldsymbol{v}). \tag{3.2}$$

The first term $\mathcal{D}(u, \boldsymbol{v})$ is a data fidelity, which serves to connect the unknown variable $\boldsymbol{v}$ to the data u. The second term $\mathcal{R}(\boldsymbol{v})$ is a regularizer that enables to incorporate a priori information about the desired solution. We will illustrate possible choices for the regularizer in the subsequent section. The parameter $\alpha \in \mathbb{R}^+$ regulates the weighting between the data fidelity and the regularizer.

To estimate motion, we incorporate the optical flow constraint into the data fidelity and complete it with a norm. In general, there are many possibilities for norms of a data fidelity. In Nikolova (2002) the characteristics of several choices are discussed. However, not all of them are useful in the context of optical flow. The norm has to be in a proper function space of the operator used in the data fidelity. Suitable choices for the estimation of optical flow are the L^1 norm or the the squared L^2 norm, which yield

$$\mathcal{D}_{Lin1}(u, \boldsymbol{v}) \coloneqq \|\boldsymbol{v} \cdot \nabla u + u_t\|_1$$

for the L^1 norm or

$$\mathcal{D}_{Lin2}(u, \boldsymbol{v}) \coloneqq \frac{1}{2} \|\boldsymbol{v} \cdot \nabla u + u_t\|_2^2$$

for the squared L^2 norm.

Both norms have different properties. The decision about which one suits better strongly depends on the data and on the kind of noise corrupting the data. The squared L^2 norm yields a least-squares solution and is thus usually a good choice if the data contains Gaussian noise. However, if there are existing strong outliers in the data, the solution

gets much worse, since the influence of those outliers is extremely high in the squared L^2 norm. In contrast to that, the L^1 norm is more robust and is therefore usually the better choice for data that contains outliers. In Black (1992) the robustness of different norms for the optical flow constraint is further analyzed.

As mentioned before, the previous approximation of the optical flow constraint is only possible if the motion between two time steps is very small. However, in real applications like microscopy images the time steps between two recordings are relatively large. Therefore, the displacement in between is often too large compared to the spatial resolution of the images. In many experimental studies this problem is solved by using a coarse-to-fine warping strategy, i.e. starting the calculations on a coarse grid such that the movement is smaller than the grid size. Then the resulting velocity field is used as an initial guess for a calculation on a finer grid. This procedures has to be continued until the original grid size is reached. However, the linearized optical flow constraint is simply not fulfilled for large displacements, which is why it is also not theoretically justified to use it in combination with a warping approach. Otherwise, the solution might get stuck in a local minimum. To overcome this problem, we introduce a nonlinear approximation in the subsequent section.

3.2.2 Nonlinear Approach

An alternative approach to deal with the brightness constancy assumption is to interpret it in a nonlinear way. The brightness constancy assumption is fulfilled if an image applied to the corresponding velocity field $\boldsymbol{v}$ is equal to the subsequent image. For two consecutive images u_0 and u_1 this means:

$$u_0(\boldsymbol{v}) = u_1,$$

which is equivalent to

$$0 = u_1 - u_0(\boldsymbol{v}). \tag{3.3}$$

To embed this constraint in a variational setting as in (3.2), it remains to equip it with a norm. Similar to Section 3.2.1, the usual choices are again the L^1 norm and the squared L^2 norm. With the L^1 norm we obtain the following data fidelity:

$$\mathcal{D}_{NL1}(u_0, u_1, \boldsymbol{v}) := \|u_1 - u_0(\boldsymbol{v})\|_1 .$$

In combination with the squared L^2 norm, we obtain

$$\mathcal{D}_{NL2}(u_0, u_1, \boldsymbol{v}) \coloneqq \frac{1}{2} \|u_1 - u_0(\boldsymbol{v})\|_2^2 .$$

In the case of a squared L^2 norm, the derived nonlinear optical flow constraints can be solved by using the well-known Gauß-Newton method. This method solves a nonlinear problem $F(\boldsymbol{v}) = y$ while linearizing it in every step. Let $F'(\boldsymbol{v})$ be the partial derivative of F with respect to $\boldsymbol{v}$. The Gauß-Newton method is defined by first calculating

$$F'(\boldsymbol{v}^n)^* F'(\boldsymbol{v}^n) r^n = -F'(\boldsymbol{v}^n)^* \left(F'(\boldsymbol{v}^n) - y\right)$$

and then updating the iterate function via

$$\boldsymbol{v}^{n+1} = \boldsymbol{v}^n + r^n.$$

This method is known to converge in combination with a proper regularizer, see Blaschke et al. (1997) (cf. Bertsekas (1999) for further analysis of the Gauß-Newton method). For the minimization problem in (3.3) combined with a regularizer $\mathcal{R}(\boldsymbol{v})$, this results in minimizing

$$\frac{1}{2} \|u_0(\hat{\boldsymbol{v}}) - u_1 + (\boldsymbol{v} - \hat{\boldsymbol{v}}) \nabla u_0(\hat{\boldsymbol{v}})\|_2^2 + \alpha \mathcal{R}(\boldsymbol{v}) \tag{3.4}$$

and updating $\hat{\boldsymbol{v}}$ as an approximation of $\boldsymbol{v}$.

We also want to use an analog approach to the Gauß-Newton method for the L^1 data fidelity $\mathcal{D}_{NL1}(u_0, u_1, \boldsymbol{v})$. However, before putting the expression

$$u_0(\hat{\boldsymbol{v}}) - u_1 + (\boldsymbol{v} - \hat{\boldsymbol{v}}) \nabla u_0(\hat{\boldsymbol{v}})$$

into an L^1 norm, we need to ensure that the term $(\boldsymbol{v} - \hat{\boldsymbol{v}}) \nabla u_0(\hat{\boldsymbol{v}})$ is well-defined in L^1.

Lemma 3.1. (Well-definedness of $\nabla u(\boldsymbol{v}) w$ in $L^1([0,T] \times \Omega)$)
Let $\Omega \in \mathbb{R}^d$ for $d \in 2, 3$, $u \in C^1(\Omega)$, and $\boldsymbol{v}, w \in BV(\Omega)^d$. Then the expression $\nabla u(\boldsymbol{v}) w$ is well-defined in $L^1([0,T] \times \Omega)$.

Proof. With the Hölder inequality we obtain

$$\|\nabla u(\boldsymbol{v}) w\|_1 \leq \|\nabla u(\boldsymbol{v})\|_\infty \cdot \|w\|_1.$$

$\|w\|_1$ is bounded due to the compact embedding $BV(\Omega) \hookrightarrow L^1(\Omega)$. From the condition $u \in C^1(\Omega)$ it follows that $\nabla u(\boldsymbol{v})$ is continuous and thus bounded in $L^\infty(\Omega)$. Consequently, $\|\nabla u(\boldsymbol{v})w\|_1$ is bounded and therefore well-defined in $L^1([0,T]\times\Omega)$. □

In comparison to the L^2 problem we need an additional damping term to ensure convergence in L^1. This term reads $\frac{c}{2}\|\boldsymbol{v}-\hat{\boldsymbol{v}}\|^2$, where c is a constant such that $c > \sup_x \|H(u_0(x))\|_2$ holds, with H being the Hessian matrix. In the following, we will show that minimizing the functional

$$J_n(\boldsymbol{v}) \coloneqq \|u_0(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v}-\boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n)\|_1 + \frac{c}{2}\|\boldsymbol{v}-\boldsymbol{v}^n\|_2^2 + \alpha\mathcal{R}(\boldsymbol{v}), \tag{3.5}$$

with $v : \Omega \to \Omega$, c as described above and $\mathcal{R}$ being a convex regularization functional, ends up in finding a minimum of the nonlinear optical flow functional

$$J(\boldsymbol{v}) \coloneqq \|u_0(\boldsymbol{v}) - u_1\|_1 + \alpha\mathcal{R}(\boldsymbol{v}). \tag{3.6}$$

We start by pointing out the existence of a minimizer of $J(\boldsymbol{v})$:

Theorem 3.1. (Existence of a Minimum of J)
Let $\Omega \in \mathbb{R}^d$ for $d \in 2, 3$, $u_0 \in C^\alpha$ for $\alpha > 0$ with $u_0(\boldsymbol{v}) = 0$ for $\boldsymbol{v} \notin \Omega$, $u_1 \in L^1(\Omega)$. Then there exists a minimum of J defined by (3.6) in $BV(\Omega)^d$.

Proof. To show the existence of a minimum we use the fundamental theorem of optimization (Theorem 2.1), which requires coercivity and lower semi-continuity of the given functional. For the latter, we need strong convergence in $L^1(\Omega)$ of $u_0(\boldsymbol{v}_n)$ for $\boldsymbol{v}_n \to \boldsymbol{v}$. To prove this, we make the following calculations:

$$\begin{aligned}\int_\Omega |u_0(\boldsymbol{v}_n) - u_0(\boldsymbol{v})|\,dx &\le \int_\Omega c|\boldsymbol{v}_n - \boldsymbol{v}|^\alpha\,dx \\ &\le \tilde{c}|\boldsymbol{v}_n - \boldsymbol{v}|_1\end{aligned}$$

for c and $\tilde{c}$ being positive constants. Hence, we achieve $u_0(\boldsymbol{v}_n) \to u_0(\boldsymbol{v})$ in $L^1(\Omega)$. Together with the convexity of $\mathcal{R}(\boldsymbol{v})$ on the Banach space $BV(\Omega)$, we obtain lower semi-continuity. Moreover, we know that the regularizer is coercive in $BV(\Omega)$. Together with the boundedness of the entire functional J, this leads to a boundedness of the term $\|u_0(\hat{\boldsymbol{v}}) - u_1\|_1$. With the triangle inequality we achieve

$$\begin{aligned}c &\ge \|u_0(\hat{\boldsymbol{v}}) - u_1\|_1 \\ &\ge \|u_0(\hat{\boldsymbol{v}})\|_1 - \|u_1\|_1.\end{aligned}$$

Since u_1 is a fixed image intensity, it is naturally bounded in $L^1(\Omega)$, hence we achieve

$\|u_1\|_1 \leq c_u$, where c_u is a constant. Solving the inequality with respect to $\|u_0(\hat{\boldsymbol{v}})\|_1$ yields

$$\|u_0(\hat{\boldsymbol{v}})\|_1 \leq c + c_u.$$

From the continuity of the operator u_0, we obtain the boundedness of $\hat{\boldsymbol{v}}$ in $L^1(\Omega)$ and consequently the coercivity of the functional J. Consequently, the conditions for the fundamental theorem of optimization are fulfilled and we obtain the existence of a minimal solution. □

We continue by showing the existence of a minimum of the iterative functional J_n:

Lemma 3.2. (Existence of a Minimum of J_n)
Let $\Omega \in \mathbb{R}^d$ for $d \in 2, 3$, $u_0 \in C^\alpha$ for $\alpha > 0$, $u_1 \in L^1(\Omega)$, and $\boldsymbol{v^n} \in BV(\Omega)^d \cap L^2(\Omega)^d$. Then there exists a minimum of the functional J_n in $BV(\Omega)$, where J_n is defined in (3.5).

Proof. From the definition of the functional $J_n(\boldsymbol{v})$ it directly follows

$$J_n(\boldsymbol{v}^n) = J(\boldsymbol{v}^n).$$

Since J_n gets minimized with respect to $\boldsymbol{v}$, we know that $J_n(\boldsymbol{v}) \leq J_n(\boldsymbol{v}^n)$ holds, for possible minimizers $\boldsymbol{v}$. Consequently, we obtain

$$J_n(\boldsymbol{v}) \leq J(\boldsymbol{v}^n).$$

Since J is finite, this gives us the coercivity of the functional J_n. The lower semi-continuity follows directly from the weak* lower-semi-continuity of norms on Banach spaces. Thus, there exists a minimum of the functional J_n. □

It remains to show that the minimum of the iterative functional (3.5) converges towards a minimum of the nonlinear functional (3.6):

Theorem 3.2. (Weak* Convergence of J_n)
Let $\Omega \in \mathbb{R}^d$ for $d = 2$, $u_0 \in C^2(\Omega)$, $u_1 \in L^1(\Omega)$, and $\boldsymbol{v} \in BV(\Omega)^d$. Furthermore, let the mapping $y \mapsto \nabla u_0(y) \cdot y$ be Lipschitz continuous. Then the sequence of minimizers $\boldsymbol{v}^{n+1}$ of the functional

$$J_n(\boldsymbol{v}) \coloneqq \|u_0(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n)\|_1 + \frac{c}{2}\|\boldsymbol{v} - \boldsymbol{v}^n\|_2^2 + \alpha\mathcal{R}(\boldsymbol{v}),$$

with $c > \sup_x \|H(u_0(x))\|_2$ and H being the Hessian matrix, weakly* converges along subsequences towards a stationary point of J in $BV(\Omega)^d$, which is given by

$$\min_{\boldsymbol{v}} J(\boldsymbol{v}) := \|u_1 - u_0(\boldsymbol{v})\|_1 + \alpha\mathcal{R}(\boldsymbol{v}).$$

Proof. To prove the statement, we need to show that every sequence $\boldsymbol{v}^{n+1}$, which fulfills the optimality condition for J_n, converges towards the minimum $\boldsymbol{v}$ of J. The optimality condition for J_n at $\boldsymbol{v}^{n+1}$ is given by

$$\nabla u_0(\boldsymbol{v}^n)s^{n+1} + \alpha p^{n+1} + c(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n) = 0, \tag{3.7}$$

where

$$s^{n+1} \in \operatorname{sgn}\left(u_0(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v}^{n+1} - \boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n)\right), \qquad \text{and} \qquad p^{n+1} \in \partial\mathcal{R}(\boldsymbol{v}^{n+1}).$$

The optimality condition for J at $\boldsymbol{v}$ is given by

$$\nabla u_0(\boldsymbol{v})s + \alpha p = 0,$$

where

$$s \in \operatorname{sgn}\left(u_0(\boldsymbol{v}) - u_1\right), \qquad \text{and} \qquad p \in \partial\mathcal{R}(\boldsymbol{v}).$$

We start by showing the boundedness of $\boldsymbol{v}^n$ in $L^2(\Omega)$. For this purpose, we use the Taylor expansion of $u_0(\boldsymbol{v}^{n+1})$ at $\boldsymbol{v}^{\boldsymbol{n}}$:

$$u_0(\boldsymbol{v}^{n+1}) = u_0(\boldsymbol{v}^n) + (\boldsymbol{v}^{n+1} - \boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n) + \frac{1}{2}(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n)H\left(u_0(\tilde{\boldsymbol{v}}^n)\right)(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n).$$

Rearranging the above equation with respect to $u_0(\boldsymbol{v}^n)$ and inserting the result into

$$J_n(\boldsymbol{v}^{n+1}) = \left\|u_0(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v}^{n+1} - \boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n)\right\|_1 + \frac{c}{2}\left\|\boldsymbol{v}^{n+1} - \boldsymbol{v}^n\right\|_2^2 + \alpha\mathcal{R}\left(\boldsymbol{v}^{n+1}\right)$$

yields

$$\begin{aligned} J_n(\boldsymbol{v}^{n+1}) = &\left\|u_0(\boldsymbol{v}^{n+1}) - u_1 - \frac{1}{2}(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n)H\left(u_0(\tilde{\boldsymbol{v}}^n)\right)(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n)\right\|_1 \\ &+ \frac{c}{2}\left\|\boldsymbol{v}^{n+1} - \boldsymbol{v}^n\right\|_2^2 + \alpha\mathcal{R}(\boldsymbol{v}^{n+1}). \end{aligned}$$

Thanks to the intermediate value theorem (cf. Forster (2006)), we know that is holds

$\tilde{\boldsymbol{v}}^n(x) \in [\boldsymbol{v}^n(x), \boldsymbol{v}^{n+1}(x)]$. Together with the Hölder inequality, this gives us

$$J_n(\boldsymbol{v}^{n+1}) \geq \left\|u_0(\boldsymbol{v}^{n+1}) - u_1\right\|_1 - \frac{1}{2}\sup_x \|H(u_0(\boldsymbol{v}^n))\|_2 \sum_{i=1}^{d} \left\|(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n)\right\|_2^2 + \frac{c}{2}\left\|\boldsymbol{v}^{n+1} - \boldsymbol{v}^n\right\|_2^2 + \alpha\mathcal{R}(\boldsymbol{v}^{n+1}).$$

Due to $c > \sup_x \|H(u_0(x))\|_2$, it follows

$$\begin{aligned} J_n(\boldsymbol{v}^n) &\geq J_n(\boldsymbol{v}^{n+1}) \\ &\geq \left\|u_0(\boldsymbol{v}^{n+1}) - u_1\right\|_1 + \alpha\mathcal{R}(\boldsymbol{v}^{n+1}) + \left(c - \sup_x \|H(u_0(\boldsymbol{v}^n))\|_2\right) \sum_{i=1}^{d} \left\|\boldsymbol{v}^{n+1} - \boldsymbol{v}^n\right\|_2^2 \\ &= J(\boldsymbol{v}^{n+1}) + \left(c - \sup_x \|H(u_0(\boldsymbol{v}^n))\|_2\right) \sum_{i=1}^{d} \left\|\boldsymbol{v}^{n+1} - \boldsymbol{v}^n\right\|_2^2. \end{aligned}$$

Since it holds $J_n(\boldsymbol{v}^n) < \infty$ and $J(\boldsymbol{v}^{n+1}) < \infty$, we achieve

$$\left(c - \sup_x \|H(u_0(\boldsymbol{v}^n))\|_2\right) \sum_{i=1}^{d} \left\|\boldsymbol{v}^{n+1} - \boldsymbol{v}^n\right\|_2^2 < \infty$$

and therefore

$$\|\boldsymbol{v}^{n+1} - \boldsymbol{v}^n\|_2^2 \to 0 \qquad \text{for } n \to \infty. \tag{3.8}$$

Thanks to the compact embedding of $BV(\Omega)$ in $L^p(\Omega)$ for $p < 2$, $\boldsymbol{v}^n$ has a weakly convergent subsequence $\boldsymbol{v}^{n_k}$ with $\boldsymbol{v}^{n_k} \to \boldsymbol{v}$ for $k \to \infty$ in $L^1(\Omega)^d$. Together with (3.8), we obtain $\boldsymbol{v}^{n_k-1} \to \boldsymbol{v}$ for $k \to \infty$ in $L^1(\Omega)^d$. For the subsequence $\boldsymbol{v}^{n_k}$ the optimality condition of J_{n_k-1} reads

$$\nabla u_0(\boldsymbol{v}^{n_k-1})s^{n_k} + \alpha p^{n_k} + c(\boldsymbol{v}^{n_k} - \boldsymbol{v}^{n_k-1}) = 0.$$

With (3.8), for the last term it holds

$$c(\boldsymbol{v}^{n_k} - \boldsymbol{v}^{n_k-1}) \to 0 \qquad \text{for } n \to \infty \text{ in } L^2(\Omega)^d.$$

The strong convergence of $\boldsymbol{v}^{n_k-1}$ in $L^1(\Omega)^d$ leads to the strong convergence

$$\nabla u_0(\boldsymbol{v}^{n_k-1}) \to \nabla u_0(\boldsymbol{v}^n) \qquad \text{for } k \to \infty \text{ in } L^1(\Omega).$$

Since it holds $-1 \leq s^{n_k} \leq 1$, we obtain that s^{n_k} is uniformly bounded in $L^\infty(\Omega)$. This gives us the weak* convergence $s^{n_k} \rightharpoonup^* s$ for $k, n \to \infty$ in $L^\infty(\Omega)$. It remains to show

the weak convergence of p^{n_k}, i.e. $p^{n_k} \rightharpoonup p$ for $k \to \infty$, to pass to the limit in (3.7). For the argument in p we start by solving the optimality condition (3.7) with respect to p^{n+1}:

$$p^{n+1} = -\frac{1}{\alpha}\left(c(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n) + \nabla u_0(\boldsymbol{v}^n)s^{n+1}\right). \tag{3.9}$$

Again, we use (3.8) to obtain

$$c(\boldsymbol{v}^{n+1} - \boldsymbol{v}^n) \to 0 \qquad \text{for } n \to \infty \text{ in } L^2(\Omega)^d.$$

Moreover, as a consequence of $u_0 \in C^2(\Omega)$, the expression $\nabla u_0(\boldsymbol{v}^n)$ is Lipschitz continuous with a Lipschitz constant that we denote by $L_{\nabla u_0}$, i.e.

$$\int |\nabla u_0(\boldsymbol{v}) - \nabla u_0(\boldsymbol{v}^n)| \leq L_{\nabla u_0} \int |\boldsymbol{v} - \boldsymbol{v}^n|.$$

Consequently, for the subsequence $\nabla u_0(\boldsymbol{v}^{n-1})$ it holds

$$\nabla u_0(\boldsymbol{v}^{n-1}) \to \nabla u_0(\boldsymbol{v}) \qquad \text{for } n \to \infty \text{ in } L^1(\Omega).$$

Due to $-1 \leq s^{n+1} \leq 1$, we additionally achieve that s^{n+1} is uniformly bounded in $L^\infty(\Omega)$. This gives us the weak* convergence $s^{n+1} \rightharpoonup^* s$ for $n \to \infty$ in $L^\infty(\Omega)$ and thus

$$\nabla u_0(\boldsymbol{v}^n)s^{n+1} \rightharpoonup \nabla u_0(\boldsymbol{v})s \qquad \text{for } n \to \infty \text{ in } L^1(\Omega).$$

Hence, p^{n+1} is bounded and weakly convergent in $L^2(\Omega)$. Since it holds $p^{n_k} \in \partial\mathcal{R}(\boldsymbol{v}^{n_k})$, it follows

$$\langle p^{n_k}, \boldsymbol{v}^{n_k}\rangle \geq \mathcal{R}(\boldsymbol{v}^{n_k})$$

for every subsequence $\boldsymbol{v}^{n_k}$ of $\boldsymbol{v}^n$. Now we need to show that the limit of $\langle p^{n_k+1}, \boldsymbol{v}^{n_k+1}\rangle$ for $k \to \infty$ exists. Inserting (3.9) into the scalar product yields

$$\begin{aligned}
\langle p^{n_k+1}, \boldsymbol{v}^{n_k+1}\rangle &= \left\langle -\frac{1}{\alpha}\left(c(\boldsymbol{v}^{n_k+1} - \boldsymbol{v}^{n_k}) + \nabla u_0(\boldsymbol{v}^{n_k})s^{n_k+1}\right), \boldsymbol{v}^{n_k+1}\right\rangle \\
&= -\frac{1}{\alpha}\left\langle \left(c(\boldsymbol{v}^{n_k+1} - \boldsymbol{v}^{n_k})\right), \boldsymbol{v}^{n_k+1}\right\rangle \\
&\quad -\frac{1}{\alpha}\int_\Omega \nabla u_0(\boldsymbol{v}^{n_k})(\boldsymbol{v}^{n_k+1} - \boldsymbol{v}^{n_k})s^{n_k+1}\,dx \\
&\quad -\frac{1}{\alpha}\int_\Omega \nabla u_0(\boldsymbol{v}^{n_k})\boldsymbol{v}^{n_k}s^{n_k+1}\,dx.
\end{aligned} \tag{3.10}$$

Due to (3.8) and the boundedness of $\nabla u_0(\boldsymbol{v}^{n_k})$ and s^{n_k+1}, the first and the second term strongly converge towards zero in $L^2(\Omega)$. In the third term, the expression $\nabla u_0(\boldsymbol{v}^{n_k})\boldsymbol{v}^{n_k}$ is Lipschitz continuous. Together with $s^{n_k+1} \rightharpoonup^* s$ for $k, n \to \infty$ in $L^\infty(\Omega)$, we obtain

$$-\frac{1}{\alpha}\int_\Omega \nabla u_0(\boldsymbol{v}^{n_k})\boldsymbol{v}^{n_k} s^{n_k+1}\,dx \to -\frac{1}{\alpha}\int_\Omega \nabla u_0(\boldsymbol{v})\boldsymbol{v}s \qquad \text{for } k \to \infty.$$

Since $\boldsymbol{v}^{n_k}$ strongly converges towards $\boldsymbol{v}$ in $L^2(\Omega)^d$, the limit

$$\lim_{k\to\infty}\langle p^{n_k+1}, \boldsymbol{v}^{n_k+1}\rangle = \langle p, \boldsymbol{v}\rangle$$

exists and it holds

$$\langle p, \boldsymbol{v}\rangle \geq \liminf_{\boldsymbol{v}} \mathcal{R}(\boldsymbol{v}^{n_k+1}) \geq \mathcal{R}(\boldsymbol{v}).$$

From $\boldsymbol{v}^{n_k+1} \rightharpoonup \boldsymbol{v}$ in $BV(\Omega)^d$, it follows $\boldsymbol{v}^{n_k+1} \to \boldsymbol{v}$ in $L^1(\Omega)^d$. Therefore, we obtain $p^{n_k} \rightharpoonup p$ in $L^2(\Omega)$, with $p \in \partial\mathcal{R}(\boldsymbol{v})$. Hence, the optimality condition is fulfilled and $\boldsymbol{v}$ is a stationary point of J. □

Consequently, we are able to minimize the iterative functional $J_n(\boldsymbol{v})$ in order to find a minimum of the nonlinear functional $J(\boldsymbol{v})$.

Remark 3.1. The embedding $BV(\Omega) \hookrightarrow L^p(\Omega)$ is only valid if $\frac{p}{p-1} \geq d$, where d is the dimension of Ω, cf. Lemma 2.3. For $d = 3$ the space $BV(\Omega)$ is not embedded in $L^2(\Omega)$. Hence, we do not obtain boundedness of $\boldsymbol{v}^{n_k+1}$ in $L^2(\Omega)$. Consequently, the scalar product in (3.10) is not necessarily converging in $L^2(\Omega)$ and thus the previously shown proof is not valid for $d = 3$.

3.3 Regularization

In this section we discuss different regularizers to complete the variational approach. In the first part of this section we present the most common regularizers that are typically used in combination with the optical flow constraint. Since those are however not perfectly fitting to many real data applications, we introduce extended regularizers in the second part. These extensions are supposed to combine wanted characteristics of the previously shown standard regularizers.

3.3.1 Standard Regularization

One of the first well-known models to estimate optical flow is the Horn-Schunck method, which was already established in 1981 in Horn and Schunck (1981). Using the

optical flow constraint in the squared L^2 norm, Horn and Schunck completed the model by adding the squared L^2 norm of the gradient of the velocity field v, i.e.

$$\mathcal{R}_{L^2}(\boldsymbol{v}) := \frac{1}{2} \|\nabla \boldsymbol{v}\|_2^2 . \tag{3.11}$$

The penalization of this term prevents the general appearance of discontinuities in the result. The estimated velocity field is smooth and the transition between different velocities or directions of movement appears without edges. A crucial quality of the squared L^2 norm is that this avoidance of discontinuities naturally circumvents the existence of noise in the velocity field. However, in many cases some edges are expected to be visible in the results, e.g. due to rigid moving objects in front of a background. These edges do not appear with this regularizer.

A variant that became more and more popular in recent years is the L^1 data fidelity combined with TV of $\boldsymbol{v}$, see e.g. Zach et al. (2007). The formal definition of TV is given by

$$TV(\boldsymbol{v}) = \sup \left\{ \int_\Omega \boldsymbol{v} \operatorname{div} \varphi \middle| \varphi \in \mathcal{C}_C^1(\Omega, Sym^1(\mathbb{R}^d)), \|\varphi\|_\infty \leq 1 \right\}.$$

Using this norm it follows:

$$\mathcal{R}_{TV}(\boldsymbol{v}) := TV(\boldsymbol{v}). \tag{3.12}$$

In contrast to (3.11), the regularization of this model penalizes the gradient of $\boldsymbol{v}$ with the L^1 norm instead of the squared L^2 norm. This leads to sparse discontinuities in the velocity field. The result typically has some strong edges, whereas the spaces between are constant. The whole velocity field thus has a cartoon-like character. Due to the strong edges, a rigid moving object is clearly distinguishable from a stationary background. This way, TV regularization allows to track moving objects like cells in microscopy data. Similar to the previous model, the preference of constant parts also avoids to obtain noise in the result. Nevertheless, this regularizer is not suited for every kind of data. TV regularization does not allow to estimate smooth transitions between velocities or directions, which appear e.g. in fluids or other elastic materials.

Both of the above mentioned regularizers can be combined with the L^1 norm as well as with the squared L^2 norm for the data fidelity, even though the previously mentioned combinations are the most popular ones. The choice of the data fidelity is mainly dependent on the noise of the images, whereas the choice of the regularizer is mainly

dependent on the type of movement that is expected. Both characteristics are not necessarily connected. However, it is not always possible to decide whether a movement between two images is either smooth, or piecewise constant. Especially in real data we often expect to find both edges and smooth parts. None of the introduced regularizers is able to estimate this combination. For this reason, we extend the regularization term in the following section.

3.3.2 Extended Regularization

In general, it is possible to use different approaches in order to achieve smooth parts in an image, e.g. higher order regularization or, as we have seen before, the application of the squared L^2 norm. In contrast to that, sharp edges are mainly achieved by using a TV-based regularizer. For this reason, we use TV as a first regularizer and introduce a second regularization term $\mathcal{R}_2(w)$ that enables smoothness. The variable $w : \mathbb{R}^d \to \mathbb{R}^d$, where d is the spatial dimension of the images, serves to link the regularization terms. To realize the linkage, we also incorporate w in the first TV regularizer and penalize the distance between the gradient of $\boldsymbol{v}$ and w and minimizing this term with respect to $\boldsymbol{v}$ and w, i.e.

$$\mathcal{R}_1(\boldsymbol{v}, w) = \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 .$$

The general variational model with extended regularization has the following structure:

$$\min_{v,w} \mathcal{D}(u, \boldsymbol{v}) + \alpha_1 \mathcal{R}_1(\boldsymbol{v}, w) + \alpha_2 \mathcal{R}_2(w).$$

The parameters $\alpha_1, \alpha_2 \in \mathbb{R}^+$ regulate both the weighting between data fidelity and regularization, as well as between the single regularization terms $\mathcal{R}_1(\boldsymbol{v}, w)$ and $\mathcal{R}_2(w)$. In our case the latter serves to adjust the proportion between smooth parts and edges. If the ratio $\frac{\alpha_2}{\alpha_1}$ is small, the number of edges is relatively large in the results. If $\frac{\alpha_2}{\alpha_1}$ is large, the number of edges is relatively small and the smooth parts predominate. The model with extended regularization that we will focus on in the remaining section reads

Model 3.1. (Extended Optical Flow Model)

$$\min_{v,w} \mathcal{D}(u, \boldsymbol{v}) + \alpha_1 \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \alpha_2 \mathcal{R}_2(w)$$

with $\mathcal{D} \in (\mathcal{D}_{Lin1}, \mathcal{D}_{Lin2}, \mathcal{D}_{NL1}, \mathcal{D}_{NL2})$.

Using this model, the velocity field v is regularized by a functional

$$\widetilde{\mathcal{R}}(\boldsymbol{v}) = \inf_{w} \left\{ \mathcal{R}_1(\boldsymbol{v}, w) + \frac{\alpha_2}{\alpha_1} \mathcal{R}_2(w) \right\}.$$

On the one hand, minimizing $\mathcal{R}(\boldsymbol{v})$ with respect to $\boldsymbol{v}$ achieves a simple TV regularization shifted by a constant depending on w. On the other hand, minimizing $\mathcal{R}_1(\boldsymbol{v}, w)$ with respect to w, forces w to be close to the TV solution of $\boldsymbol{v}$. Consequently, the second regularization term $\mathcal{R}_2(w)$ is applied to a solution that already approximates a TV solution.

The first approach to achieve both characteristics, sharp edges and smooth parts, is to combine those regularizers that we already introduced in Section 3.3.1. We previously discussed that the desired characteristics are accessible with one of them at a time. TV is already incorporated in Model 3.1. Thus, we use the squared L^2 norm in the remaining term $\mathcal{R}_2(w)$. Since w already approximates the gradient of $\boldsymbol{v}$, we do not need another gradient for w in $\mathcal{R}_2(w)$ in order to enable smoothness. Consequently, using

$$\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w) = \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \frac{\alpha_2}{2\alpha_1} \sum_{i=1}^{d} \|w_i\|_2^2, \tag{3.13}$$

we obtain the following model:

Model 3.2. (TV/L^2 Optical Flow Model)

$$\min_{v,w} \mathcal{D}(u, \boldsymbol{v}) + \alpha_1 \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \frac{\alpha_2}{2} \sum_{i=1}^{d} \|w_i\|_2^2$$

We are able to show that $\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w)$ is Lipschitz equivalent to $TV(\boldsymbol{v})$:

Lemma 3.3. (Lipschitz Equivalence of $\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w)$ and $TV(\boldsymbol{v})$)
Let $\boldsymbol{v} \in L^1(\Omega)$. Then there exist constants $0 < c < C < \infty$ such that it holds

$$cTV(\boldsymbol{v}) \leq \mathcal{R}_{TV/L^2}(\boldsymbol{v}, w) \leq CTV(\boldsymbol{v}).$$

Proof. Let $\boldsymbol{v} \in BV(\Omega)$. For every $w \in L^2(\Omega)$ we obtain

$$\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w) \leq \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \frac{\alpha_2}{2\alpha_1} \sum_{i=1}^{d} \|w_i\|_2^2.$$

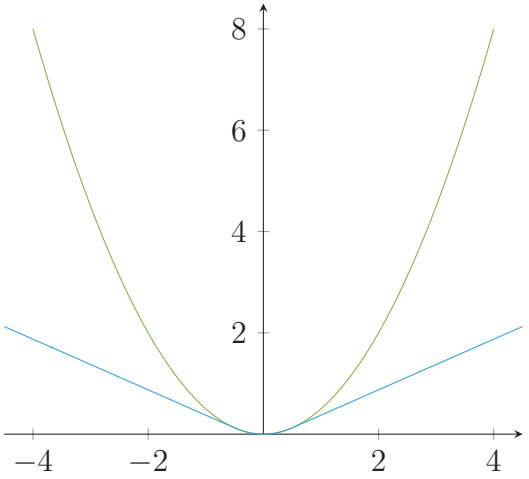

Figure 3.3: Blue: Huber loss function for $\delta = 0.5$, green: $f(x) = \frac{x^2}{2}$.

Regarding $w_i = 0$ for $i = 1, ..., d$ gives us

$$\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w) \leq \sum_{i=1}^{d} \|\nabla v_i\|_1 ,$$

and thus for $C = \alpha_1$ it holds $\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w) \leq CTV(\boldsymbol{v})$.
For the other direction, let $w \in L^2$. The triangle inequality gives us

$$\sum_{i=1}^{d} \|\nabla v_i\|_1 \leq \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \sum_{i=1}^{d} \|w_i\|_1 .$$

Due to $L^2(\Omega) \subset L^1(\Omega)$, we obtain

$$\sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \sum_{i=1}^{d} \|w_i\|_1 \leq \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + |\Omega|^{\frac{1}{2}} \sum_{i=1}^{d} \|w_i\|_2 .$$

Let $c_0 = d \cdot \max_i(\|w_i\|_2)$ for $i = 1, ..., d$. With $c_1 = \max(1, |\Omega|^{\frac{1}{2}}/(\frac{\alpha_2}{2\alpha_1} c_0))$ we achieve

$$\sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + |\Omega|^{\frac{1}{2}} \sum_{i=1}^{d} \|w_i\|_2 \leq c_1 \left(\sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \frac{\alpha_2}{2\alpha_1} \sum_{i=1}^{d} \|w_i\|_2^2 \right),$$

and consequently for $c = \frac{1}{c_1}$ it holds $cTV(\boldsymbol{v}) \leq \mathcal{R}_{TV/L^2}(\boldsymbol{v}, w)$. □

With the above lemma we are able to adopt existence and uniqueness results for solutions from known results for $BV(\Omega)$.

For $|\nabla v| = x$ and $\frac{\alpha_2}{\alpha_1} = \delta$, the minimum of $\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w)$ with respect to w is equivalent to the well-known Huber loss function, see Huber et al. (1964) and Huber (1973),

which is defined as follows:

$$H_\delta(x) := \begin{cases} \frac{x^2}{2} & |x| \leq \delta \\ \delta|x| - \frac{\delta^2}{2} & |x| > \delta \end{cases}. \tag{3.14}$$

Figure 3.3 shows the Huber loss function compared to a fully quadratic function. We prove this statement with the following theorem:

Theorem 3.3. (Equivalence of TV/L^2 Regularizer and Huber norm)
The functional $F : BV(\Omega) \to \mathbb{R}$ with

$$F(\boldsymbol{v}) = \min_w \alpha_1 \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \frac{\alpha_2}{2} \sum_{i=1}^{d} \|w_i\|_2^2 \tag{3.15}$$

and $\alpha_1, \alpha_2 > 0$, is equivalent to

$$F(\boldsymbol{v}) = \alpha_1 \sum_{i=1}^{d} \int_\Omega H_{\frac{\alpha_2}{\alpha_1}}(\nabla v_i)\, dx,$$

where $H : \mathbb{R}^d \to \mathbb{R}$ is defined in (3.14).

Proof. Using the definition of the L^1 and L^2 norms, we write $F(\boldsymbol{v})$ as

$$F(\boldsymbol{v}) = \alpha_1 \left(\min_w \sum_{i=1}^{d} \int_\Omega |\nabla v_i - w_i| + \frac{\alpha_2}{2\alpha_1} |w_i|^2 \, dx \right).$$

For $i = \{1, ..., d\}$ we calculate the optimality condition for the subproblem

$$\min_{w_i} |\nabla v_i - w_i| + \frac{\alpha_2}{2\alpha_1} |w_i|^2. \tag{3.16}$$

Since the L^1 norm is not Fréchet differentiable in $\nabla v_i - w_i = 0$, we need to distinguish between the cases $w_i < \nabla v_i$, $w_i > \nabla v_i$, and $w_i = \nabla v_i$.
First case: $w_i < \nabla v_i$
Since the inside of the L^1 norm is positive in this case, the optimality condition of (3.16) reads

$$-1 + \frac{\alpha_2}{\alpha_1} w_i \stackrel{!}{=} 0,$$

which is equivalent to $w_i = \frac{\alpha_1}{\alpha_2}$. Inserting this condition into $w_i < \nabla v_i$ shows that this first case occurs for $\nabla v_i > \frac{\alpha_1}{\alpha_2}$.
Second case: $w_i > \nabla v_i$
Equivalent calculations to the first case result in $w_i = -\frac{\alpha_1}{\alpha_2}$, for $\nabla v_i < \frac{\alpha_1}{\alpha_2}$.

Third case: $w_i = \nabla v_i$
Per definition we know $w_i = \nabla v_i$. Concerning the condition on w_i, this case covers the remaining part, i.e. $|\nabla v_i| \leq \frac{\alpha_1}{\alpha_2}$.
Combining these three cases we obtain

$$\arg\min_{w_i} \left\{ |\nabla v_i - w_i| + \frac{\alpha_2}{2\alpha_1}|w_i|^2 \right\} = \begin{cases} \frac{\alpha_1}{\alpha_2}\frac{\nabla v_i}{|\nabla v_i|} & \text{if } |\nabla v_i| > \frac{\alpha_1}{\alpha_2} \\ \nabla v_i & \text{if } |\nabla v_i| \leq \frac{\alpha_1}{\alpha_2} \end{cases}.$$

We insert this obtained minimum into (3.16) and yield

$$\min_{w_i} |\nabla v_i - w_i| + \frac{\alpha_2}{2\alpha_1}|w_i|^2 = \begin{cases} |\nabla v_i| - \frac{\alpha_1}{2\alpha_2} & \text{if } |\nabla v_i| > \frac{\alpha_1}{\alpha_2} \\ \frac{\alpha_2}{2\alpha_1}|\nabla v_i|^2 & \text{if } |\nabla v_i| \leq \frac{\alpha_1}{\alpha_2} \end{cases},$$

which is exactly the definition of $H_{\frac{\alpha_2}{\alpha_1}}(\nabla v_i)$. Embedding this result for the subproblem into the functional (3.15), we end up with

$$F(\boldsymbol{v}) = \alpha_1 \sum_{i=1}^{d} \int_\Omega H_{\frac{\alpha_2}{\alpha_1}}(\nabla v_i)\, dx.$$

□

In Burger et al. (2016b) a more general version of this regularizer with arbitrary L^p norm for the second term is studied. A variational model for optical flow with an anisotropic Huber regularization was e.g. analyzed in Werlberger et al. (2009).

A method to obtain smoothness without using the squared L^2 norm is the application of higher order derivatives. We achieve this by using the L^1 norm of the gradient of w for the second regularization term, which yields

$$\mathcal{R}_{TV/TV}(\boldsymbol{v}, w) = \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \frac{\alpha_2}{2\alpha_1} \sum_{i=1}^{d} \|\nabla w_i\|_1. \tag{3.17}$$

Altogether, we achieve

Model 3.3. (TV/TV Optical Flow Model)

$$\min_{v,w} \mathcal{D}(u, \boldsymbol{v}) + \alpha_1 \sum_{i=1}^{d} \|\nabla v_i - w_i\|_1 + \frac{\alpha_2}{2} \sum_{i=1}^{d} \|\nabla w_i\|_1.$$

Even though there is only a simple gradient in the second term, we receive a second order

derivative due to the fact that w already approximates the gradient of $\boldsymbol{v}$. The gradient of w consequently approximates the second order derivative of $\boldsymbol{v}$. While the first order derivative, which builds the standard TV regularizer, is zero only for constant parts, the second order derivative is zero for all kinds of linear functionals. Therefore, the combination of first and second order derivatives allows to reconstruct affine linear functionals.

The regularizer (3.17) is equivalent to a primal realization of a total generalized variation (TGV) regularizer, which was established in Bredies et al. (2010). The general definition of TGV is

$$TGV_{\alpha}^{k}(eq:tvtv) = sup\left\{\int_{\Omega} eq:tvtv \,\mathrm{div}^{k}\varphi \, dx \;\middle|\; \varphi \in \mathcal{C}_{C}^{k}(\Omega, Sym^{k}(\mathbb{R}^{d})), \right. \\ \left. \|div^{l}\varphi\|_{\infty} \leq \alpha_{l},\, l = 0, ..., k-1\right\}.$$

While TGV was mainly used for denoising studies in the beginning, recently it got more popular for motion estimation problems as well, see e.g. Ranftl et al. (2012) and Ranftl et al. (2014). For $k = 2$ this regularizer is equivalent to (3.17). This fact allows us to use known results about the TGV regularizer:

Lemma 3.4. (Properties of $\mathcal{R}_{TV/TV}(\boldsymbol{v}, w)$)
$\mathcal{R}_{TV/TV}(\boldsymbol{v}, w)$ is proper, convex and lower semi-continuous on $L^{p}(\Omega)$, for $1 \leq p \leq \infty$.

Proof. A proof for the equivalent TGV regularizer can be found in Bredies et al. (2010). □

Similar to $\mathcal{R}_{TV/TV}(\boldsymbol{v}, w)$, it is also possible to show Lipschitz equivalence between $TV(\boldsymbol{v})$ and $\mathcal{R}_{TV/TV}(\boldsymbol{v}, w)$ in combination with $\|\boldsymbol{v}\|_{1}$:

Lemma 3.5. (Lipschitz Equivalence of $\|\boldsymbol{v}\|_{1} + \mathcal{R}_{TV/TV}(\boldsymbol{v}, w)$ and $TV(\boldsymbol{v})$)
Let $\boldsymbol{v} \in L^{1}(\Omega)$. Then there exist constants $0 < c < C < \infty$ such that it holds

$$cTV(\boldsymbol{v}) \leq \|\boldsymbol{v}\|_{1} + \mathcal{R}_{TV/TV}(\boldsymbol{v}, w) \leq CTV(\boldsymbol{v}).$$

Proof. A proof for the equivalent TGV regularizer can be found in Bredies and Valkonen (2011). □

3.4 Bregman Iterations

TV regularization is known to have a typical failure, which is the loss of contrast. This loss happens for all kinds of data fidelities. In reconstruction approaches, the intensities in the results are lower than in ground truth images. In motion estimation approaches,

the velocities in the results are lower than the actual ones. Unfortunately, this loss of contrast does not only appear with the standard TV regularizer, but also with other regularizers that contain TV, like those we introduced in Section 3.3.2.

For squared L^2 data fidelities, a possibility to overcome this problem is the use of Bregman iterations, as they were proposed for linear problems in Osher et al. (2005). A generalization to L^1 data fidelities will be mentioned at the end of this section. An analysis of Bregman iterations for TV-based regularizers can be found e.g. in Burger and Osher (2013), Osher et al. (2005) and Burger et al. (2007b). The idea of this method is to use a Bregman distance instead of the usual regularization term. For $F : \Omega \to \mathbb{R}$ and $a, b \in \Omega$, the generalized Bregman distance is defined as

$$D_F^b(p, q) = F(p) - F(q) - \langle b, p - q\rangle, \qquad \text{with } b \in \partial F(q),$$

where b is the Bregman variable. For the Bregman iterations, the specific distance term $D_{\mathcal{R}}^{b^n}(\boldsymbol{v}, \boldsymbol{v}^n)$ depends on the choice of the regularizer. For the standard TV regularizer (3.12), the distance term has the following appearance:

$$D_{\mathcal{R}}^{\boldsymbol{b}^n}(\boldsymbol{v}, \boldsymbol{v}^n) = \mathcal{R}(\boldsymbol{v}) - \mathcal{R}(\boldsymbol{v}^n) - \langle \boldsymbol{b}^n, \boldsymbol{v} - \boldsymbol{v}^n\rangle, \qquad \text{with } \boldsymbol{b}^n \in \partial\mathcal{R}(\boldsymbol{v}^n).$$

For the extended regularizers (3.13) and (3.17), it appears as

$$D_{\mathcal{R}}^{\boldsymbol{b}^n}(\boldsymbol{v}, \boldsymbol{v}^*) = \mathcal{R}(\boldsymbol{v}, w) - \mathcal{R}(\boldsymbol{v}^n, w) - \langle \boldsymbol{b}^n, \boldsymbol{v} - \boldsymbol{v}^n\rangle, \qquad \text{with } \boldsymbol{b}^n \in \partial\mathcal{R}(\boldsymbol{v}^n, w).$$

Basically, Bregman iterations consist of two major steps, which are calculating the main variable of the particular minimization problem, and calculating the variable of the associated Bregman distance. These results are achieved by using an iterative approximation. For the illustration of the Bregman iterations, we consider a general inverse problem

$$Av = g, \tag{3.18}$$

where $A : L^2(\Omega) \to \mathcal{H}$ is a bounded, linear operator whose kernel does not include the space of continuous functions, and $\mathcal{H}$ is a Hilbert space. Instead of solving the regularized minimization problem

$$\min_v \frac{1}{2}\|Av - g\|^2 + \alpha\mathcal{R}(v), \tag{3.19}$$

the Bregman iterations are defined as follows:

Definition 3.1. (Bregman Iterations for Linear Problems)
Let A be as in (3.18). For $v_0 = b_0 = 0$ and for $n = 1, 2, \cdots$ compute

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \frac{1}{2}\|A\boldsymbol{v} - g\|^2 + \alpha D_{\mathcal{R}}^{\boldsymbol{b}^n}(\boldsymbol{v}, \boldsymbol{v}^n), \tag{3.20}$$

$$\boldsymbol{b}^{n+1} = \boldsymbol{b}^n + A^*(g - A\boldsymbol{v}^{n+1}). \tag{3.21}$$

The first step (3.20) of a Bregman iteration is the update of the main variable. For this purpose, the regularizer $\mathcal{R}(\boldsymbol{v})$ in the general minimization problem (3.19) is replaced by the Bregman distance regarding $\mathcal{R}$ between the minimizer $\boldsymbol{v}$ and the previous iterate $\boldsymbol{v}^n$. The second step (3.21) is the update of the Bregman variable. It has the effect of adding back the contrast error that occurred after a couple of iterations of the main algorithm. The fact that $\boldsymbol{b}$ is always in the subgradient of the regularizer ensures that all edges that occur in the results remain at their correct position. Nevertheless, we need to mention that it is not useful to calculate too many Bregman iterations. After a couple of iterations the frequent addition of the calculated error leads to a vanishing effect of the original minimization functional. As a consequence, e.g. in a denoising approach the noise returns after a distinct amount of Bregman iterations. Without stopping at a certain point the iterates would converge towards the original noisy image. In Osher et al. (2005) it is proposed to stop the iterations as soon as the new iterate $\boldsymbol{v}^{n+1}$ is more noisy than the previous $\boldsymbol{v}^n$. The amount of necessary Bregman iterations is coupled to the choice of the regularization parameter α, which serves as a step length for the Bregman iterations.

The term $\langle \boldsymbol{b}^n, \boldsymbol{v}^n \rangle$, which is part of

$$\langle \boldsymbol{b}^n, \boldsymbol{v} - \boldsymbol{v}^n \rangle = \langle \boldsymbol{b}^n, \boldsymbol{v} \rangle - \langle \boldsymbol{b}^n, \boldsymbol{v}^n \rangle,$$

as well as the term $\mathcal{R}(\boldsymbol{v}^n)$ are not dependent on $\boldsymbol{v}$. Considering this, we shorten (3.20) by writing

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \frac{1}{2}\|A\boldsymbol{v} - g\|^2 + \alpha\mathcal{R}(\boldsymbol{v}) - \langle \boldsymbol{b}^n, \boldsymbol{v} \rangle. \tag{3.22}$$

With this formulation we see that the Bregman iterations are in fact equivalent to the well known augmented Lagrangian method (see (Rockafellar 1976)), which is

defined as follows:

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \frac{1}{2}\|A\boldsymbol{v} - g\|^2 + \alpha\mathcal{R}(\boldsymbol{v}) + \langle \boldsymbol{b}^n_{ALM}, g - A\boldsymbol{v}\rangle, \tag{3.23}$$

$$\boldsymbol{b}^{n+1}_{ALM} = \boldsymbol{b}^n_{ALM} + g - A\boldsymbol{v}^{n+1}. \tag{3.24}$$

To show the equivalence, we note that it holds

$$\boldsymbol{b}^n = A^*\boldsymbol{b}^n.$$

Inserting this into (3.22) yields

$$\begin{aligned}\boldsymbol{v}^{n+1} &= \arg\min_{\boldsymbol{v}} \frac{1}{2}\|A\boldsymbol{v} - g\|^2 + \alpha\mathcal{R}(\boldsymbol{v}) - \langle A^*\boldsymbol{b}^n_{ALM}, \boldsymbol{v}\rangle \\ &= \arg\min_{\boldsymbol{v}} \frac{1}{2}\|A\boldsymbol{v} - g\|^2 + \alpha\mathcal{R}(\boldsymbol{v}) - \langle \boldsymbol{b}^n_{ALM}, A\boldsymbol{v}\rangle,\end{aligned}$$

which is equivalent to (3.23). It is also possible to avoid the scalar product $\langle \boldsymbol{b}^n, \boldsymbol{v}\rangle$ in (3.22). Due to the squaring of the data fidelity, the scalar product can be included in this fidelity. The problem in $\boldsymbol{v}$ then looks as follows:

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \frac{1}{2}\|A\boldsymbol{v} - (g + \boldsymbol{b}^n)\|^2 + \alpha\mathcal{R}(\boldsymbol{v}). \tag{3.25}$$

In Osher et al. (2005), there are also shown convergence results for the above system (3.20) – (3.21), where $\mathcal{R}$ is the TV norm:

Theorem 3.4. (Weak* Convergence of Bregman Iterations for Linear Problems) Assume there exists $\bar{v} \in BV(\Omega)$ with $A\bar{v} = f$. Then

$$\|Av_k - g\|^2_{\mathcal{H}} \leq k^{-1}\mathcal{R}(\bar{v}),$$

v_k has a weak* convergent subsequence in $BV(\Omega)$, and the limit of each weak* convergent subsequence is a solution of $Av = g$. If $\bar{v}$ is the unique solution, $v_{k*} \rightharpoonup^* \bar{v}$ in $BV(\Omega)$.

The proof can be found in Osher et al. (2005). To apply Bregman iterations to the linearized optical flow models, we need to use the formulation of the optical flow constraint as inverse problems. In Section 3.2.1 we showed that the linearized constraint can be written in the form $Av = g$, if it holds

$$A\boldsymbol{v} = (\nabla u)^T \cdot \boldsymbol{v} \qquad \text{and} \qquad g = -u_t.$$

Inserting this into (3.20) and (3.21) yields

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \frac{1}{2}\|u_t + \nabla u \cdot \boldsymbol{v}\|^2 + \alpha D_{\mathcal{R}}^{b^n}(\boldsymbol{v}, \boldsymbol{v}^n),$$
$$\boldsymbol{b}^{n+1} = \boldsymbol{b}^n - (u_t + \nabla u \cdot \boldsymbol{v}^n)\nabla u.$$

Since the operator $A = (\nabla u)^T$ fulfills the conditions of Theorem 3.4, the convergence result is also true for the linearized optical flow model with L^2 data fidelity and TV regularization.

In the original formulation, Bregman iterations are only defined for linear inverse problems. As already mentioned, in the case of the L^2 data fidelity, the nonlinear constraint can be regarded as the Gauß-Newton method for nonlinear minimization problems. Therefore, we introduce a generalization of Definition 3.1 for nonlinear problems, which is also based on the Gauß-Newton method. This can also be considered as a generalization of the Levenberg-Marquardt method. See Moré (1978) for further details of this algorithm. In the original formulation, the Levenberg Marquardt method consists of the Gauß-Newton method in combination with a Tikhonov regularization, which in the following generalization is replaced by the Bregman distance.

Definition 3.2. (Generalized Levenberg-Marquardt Iteration)
Let F be a nonlinear, bounded and Fréchet differentiable operator with $F(\boldsymbol{v}) = y$. For $v_0 = b_0 = 0$ and for $n = 1, 2, \cdots$ compute

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \frac{1}{2}\left\|y - F(\boldsymbol{v}^n) + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla F(\boldsymbol{v}^n)\right\|^2 + \alpha^n D_{\mathcal{R}}^{b^n}(\boldsymbol{v}, \boldsymbol{v}^n), \tag{3.26}$$
$$\boldsymbol{b}^{n+1} = \boldsymbol{b}^n - \frac{1}{\alpha^n}\left(y - F(\boldsymbol{v}^n) + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla F(\boldsymbol{v}^n)\right)\nabla F(\boldsymbol{v}^{n+1}). \tag{3.27}$$

A similar approach can also be found in Bachmayr (2007) and Bachmayr and Burger (2009). In contrast to the linear case in Definition 3.1, the regularization parameter α^n in Definition 3.2 is dependent on the current iteration step n. This property is necessary to achieve suitable convergence results.

For $F(\boldsymbol{v}) = u_1(\boldsymbol{v})$ and $y = u_0$, (3.26) gives the nonlinear optical flow constraint combined with the Bregman distance, whereas $\boldsymbol{v}^n$ can be regarded as $\hat{\boldsymbol{v}}$. Altogether, the Bregman iteration scheme for the nonlinear optical flow models with L^2 data fidelity reads:

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \frac{1}{2}\left\|u_0 - u_1(\hat{\boldsymbol{v}}) + (\boldsymbol{v} - \hat{\boldsymbol{v}})\nabla u_1(\hat{\boldsymbol{v}})\right\|^2 + \alpha^n D_{\mathcal{R}}^{b^n}(\boldsymbol{v}, \hat{\boldsymbol{v}}),$$
$$\boldsymbol{b}^{n+1} = \boldsymbol{b}^n - \frac{1}{\alpha^n}\left(u_0 - u_1(\hat{\boldsymbol{v}}) + (\boldsymbol{v} - \hat{\boldsymbol{v}})\nabla u_1(\hat{\boldsymbol{v}})\right)\nabla u_1(\hat{\boldsymbol{v}}).$$

Up to this point we only discussed possibilities for iterative algorithms including Bregman distances for squared L^2 data fidelities. Even though the iterations themselves would also be solvable for L^1 data fidelities, the well-definedness and convergence results are not transferable to non-quadratic norms. One reason for that is the fact that it is not possible to include the scalar product of the Bregman distance into an L^1 norm, as we did in (3.25). In He et al. (2006) a generalization of (3.20), which is also applicable to L^1 data fidelities, is introduced. The main idea behind this is to divide the regularization parameter by two after each iteration step. Thus, the iteration step looks as follows:

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} \|A\boldsymbol{v} - g\| + \frac{\alpha}{2^n} D_{\mathcal{R}}^{b^n}(\boldsymbol{v}, \boldsymbol{v}^n). \tag{3.28}$$

With this modification it is possible to obtain well-definedness and achieve convergence results.

3.5 Primal-Dual Formulation

To solve the previously proposed motion estimation models, we use the Chambolle-Pock algorithm introduced in Definition 2.24. To use the iteration scheme we need to assign the single terms of our models to the terms F and G in the notation of (2.1). Since we are solving the models with respect to the vector field $\boldsymbol{v}$, we assign the terms based on the presence of an arbitrary operator, which is applied to $\boldsymbol{v}$. The optical flow constraint $\boldsymbol{v} \cdot \nabla u + u_t$ contains the gradient operator ∇ and a temporal derivative, but both of them are only applied to the image intensity u and not to $\boldsymbol{v}$. Thus, we always assign the data fidelity to G. The standard regularizers from Section 3.3.1 both consist of a single term containing the gradient operator applied to $\boldsymbol{v}$ complemented with a norm. Consequently, we assign them to F. For the extended regularizers from Section 3.3.2 we need to consider the fact that these models are not only solved with regard to $\boldsymbol{v}$. The models get also minimized with respect to the auxiliary variable w. For this reason, we additionally assign those terms to F, which contain an operator applied to w. Since the extended regularizers consist of more than a single term, we regard every term independently. The TV/L^2 regularizer in Model 3.2 contains the gradient operator applied to v in the first term, but no operator at all in the second term. Therefore, we assign the first term to F and the second term to G. The TV/TV regularizer in Model 3.3 contains the gradient operator applied to $\boldsymbol{v}$ in the first term and the same operator applied to w in the second term. Hence, we assign both of them to F. Table 3.1 gives an overview of the previously described assignments. The data fidelity $\mathcal{D}$ is chosen

Regularizer	G	F	K
L^2	$\mathcal{D}(u, \boldsymbol{v})$	$\frac{\alpha}{2} \left\|\nabla \boldsymbol{v}\right\|_2^2$	∇
TV	$\mathcal{D}(u, \boldsymbol{v})$	$\alpha \left\|\nabla \boldsymbol{v}\right\|_1$	∇
TV/L^2	$\mathcal{D}(u, \boldsymbol{v}) + \frac{\alpha_1}{2} \left\|w\right\|_2^2$	$\alpha_0 \left\|\nabla \boldsymbol{v} - w\right\|_1$	$\begin{pmatrix} \nabla & -I \end{pmatrix}$
TV/TV	$\mathcal{D}(u, \boldsymbol{v})$	$\alpha_0 \left\|\nabla \boldsymbol{v} - w\right\|_1 + \alpha_1 \left\|\nabla w\right\|_1$	$\begin{pmatrix} \nabla & -I \\ 0 & \nabla \end{pmatrix}$

Table 3.1: Assignment of the single terms to G and F.

out of $(\mathcal{D}_{Lin1}, \mathcal{D}_{Lin2}, \mathcal{D}_{NL1}, \mathcal{D}_{NL2})$. The operator K in the right column of the table is a matrix that combines all occurring operators in the particular regularizer. For the extended regularizers K needs to be applied to the vector $\begin{pmatrix} \boldsymbol{v} & w \end{pmatrix}^T$ to achieve the desired model.

In the case of Bregman iterations the problem for $\boldsymbol{v}$ can also be written in the notation of (2.1). In comparison to the previously described models without Bregman iterations, we only need to consider the additional $\boldsymbol{b}^n$ in the data fidelity. This is at least true for squared L^2 data fidelities as we already showed with (3.25). Since $\boldsymbol{b}^n$ serves as a constant in this step, this term does not change the assignment to G. In combination with an L^1 norm, the modified iterations (3.28) contain an additional term in form of the scalar product $\langle \boldsymbol{b}^n, \boldsymbol{v} \rangle$, which cannot be included in the data fidelity in this case. Since there is still no operator in this term, it is also assigned to G. The calculations in the remaining parts of this section will not explicitly handle this additional term in G. However, the proceeding for the calculations of the resolvent operators is equivalent to Section 3.5.3.

Since we want to use the Chambolle-Pock framework (Definition 2.24) for solving the different problems, we need to rewrite the models using their equivalent saddle point formulations. For the models with standard regularization, the saddle point formulation is given by

$$\min_{v \in \mathcal{X}} \max_{p \in \mathcal{Y}} G(v) + \langle v, K^* p \rangle - F^*(p), \tag{3.29}$$

where p is the dual variable of v. For the L^1-TV/L^2 model, the additional variable w,

has to be included in the saddle point formulation, i.e.

$$\min_{v,w\in\mathcal{X}} \max_{p\in\mathcal{Y}} G(v,w) + \left\langle \begin{pmatrix} v \\ w \end{pmatrix}, K^* p \right\rangle - F^*(p). \tag{3.30}$$

Due to the additional term in the remaining L^1-TV/TV model, we also need to consider an additional dual variable q for the saddle point formulation:

$$\min_{v,w\in\mathcal{X}} \max_{p,q\in\mathcal{Y}} G(v) + \left\langle \begin{pmatrix} v \\ w \end{pmatrix}, K^* \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle - F^* \begin{pmatrix} p \\ q \end{pmatrix}. \tag{3.31}$$

To use these formulations, we need to calculate the adjoint operators K^* for the given operators K, as well as the convex conjugates $F^*(p)$ of $F(Kv)$ respectively $F^*(p,q)$ of $F\left(K(v,w)\right)$. For the Chambolle-Pock algorithm itself, we also need the resolvent operators for all possible terms G and F^*. We start by calculating the adjoint operators and subsequently derive the convex conjugates.

3.5.1 Adjoint Operators

As it is shown in Table 3.1, there are three different operators occurring in the presented models. The first one is $K = \nabla$. The adjoint operator of ∇ is given by $-\nabla\cdot$, and thus $K^* = -\nabla\cdot$. Therefore, the scalar product in (3.29) has the following structure:

$$\langle v, K^* p\rangle = -\langle v, \nabla \cdot p\rangle .$$

The operator $K = \begin{pmatrix} \nabla & -I \end{pmatrix}$ additionally contains the identity, of which the adjoint operator is again the identity. This leads to $K^* = -\begin{pmatrix} \nabla\cdot & I \end{pmatrix}^T$. Hence, the scalar product in (3.30) becomes

$$\left\langle \begin{pmatrix} v \\ w \end{pmatrix}, K^* p \right\rangle = -\left\langle \begin{pmatrix} v \\ w \end{pmatrix}, \begin{pmatrix} \nabla\cdot \\ I \end{pmatrix} p \right\rangle .$$

The remaining operator and its adjoint operator are given by

$$K = \begin{pmatrix} \nabla & -I \\ 0 & \nabla \end{pmatrix} \qquad \text{and} \qquad K^* = -\begin{pmatrix} \nabla\cdot & 0 \\ I & \nabla\cdot \end{pmatrix}.$$

For the scalar product in (3.31) we need to consider the additional dual variable q in

order to achieve

$$\left\langle \begin{pmatrix} v \\ w \end{pmatrix}, K^* \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle = - \left\langle \begin{pmatrix} v \\ w \end{pmatrix}, \begin{pmatrix} \nabla\cdot & 0 \\ I & \nabla\cdot \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle .$$

3.5.2 Convex Conjugates

Lemma 3.6. (Convex Conjugate for Squared L^2 Norm)
For $F(w) = \frac{\alpha}{2} \|w\|_2^2$ the convex conjugate is given by

$$F^*(p) = \frac{1}{2\alpha} \|p\|_2^2.$$

Proof. We start by defining an auxiliary functional $\widetilde{F}(w) = \frac{1}{2}\|w\|_2^2$. With the definition of the convex conjugate, we obtain:

$$\widetilde{F}^*(p) = \sup_w \left\{ \langle p, w \rangle - \frac{1}{2}\|w\|_2^2 \right\}.$$

Since the functional $\widetilde{F}^*(p)$ is Fréchet differentiable, we are able to find an optimal point $\hat{x}$ that satisfies $\hat{x} = p$. For this point we calculate the convex conjugate $\widetilde{F}^*(p)$ of the auxiliary functional $\widetilde{F}(w)$:

$$\widetilde{F}^*(p) = \langle p, p \rangle - \frac{1}{2}\|p\|_2^2 = \frac{1}{2}\|p\|_2^2.$$

For $F(w) = \alpha\widetilde{F}(w)$ we use Lemma 2.2 to identify $F^*(p)$:

$$F^*(p) = \left(\alpha\widetilde{F}(p)\right)^* = \alpha\widetilde{F}^*\left(\frac{p}{\alpha}\right) = \frac{\alpha}{2}\left\|\frac{p}{\alpha}\right\|_2^2 = \frac{1}{2\alpha}\|p\|_2^2.$$

□

Lemma 3.7. (Convex Conjugate for L^1 Norm)
For $F(w) = \alpha \|w\|_1$ the convex conjugate is given by $F^*(p) = \alpha\delta_{B(L^\infty)}\left(\frac{p}{\alpha}\right)$, where

$$\delta_{B(L^\infty)}(p) := \begin{cases} 0, & p \in B_{(L^\infty)} \\ \infty, & \text{else} \end{cases}.$$

Proof. Again, we start by defining an auxiliary functional $\widetilde{F}(w) = \|w\|_1$. We distinguish between the two cases $p \in B(L^\infty)$ and $p \notin B(L^\infty)$.

1. $p \in B(L^\infty) \Leftrightarrow \|p\|_\infty \leq 1$
   With the help of the Cauchy-Schwartz inequality, for the scalar product $\langle p, w \rangle$ we

obtain:

$$\langle p, w \rangle \leq \|p\|_\infty \|w\|_1.$$

Consequently, it also holds

$$\langle p, w \rangle \leq \|w\|_1,$$

which is equivalent to

$$\langle p, w \rangle - \|w\|_1 \leq 0. \tag{3.32}$$

Due to the definition of the convex conjugate, we achieve

$$\widetilde{F}^*(p) \geq \langle p, w \rangle - \|w\|_1, \quad \forall w \in L^1.$$

For $w = 0$ it holds $\langle p, 0 \rangle - \|0\|_1 = 0$ and consequently $\widetilde{F}^*(p) \geq 0$. Combining this with (3.32) yields $\widetilde{F}^*(p) = 0$.

2. $p \notin B(L^\infty) \Leftrightarrow \|p\|_\infty > 1$
   The definition of the norm in a dual space gives us

$$\|p\|_\infty = \sup_{w \in L^1 \setminus \{0\}} \frac{\langle p, w \rangle}{\|w\|_1}.$$

Due to $\|p\|_\infty > 1$, there exists an element $w_0 \in L^1 \setminus \{0\}$ such that

$$\frac{\langle p, w_0 \rangle}{\|w_0\|_1} > 1,$$

which is equivalent to $\langle p, w_0 \rangle > \|w_0\|_1$ and hence

$$\langle p, w_0 \rangle - \|w_0\|_1 > 0. \tag{3.33}$$

For every element $w_0 \in L^1 \setminus \{0\}$ and every constant $c > 0$, the product cw_0 is also an element of $L^1 \setminus \{0\}$. Together with the definition of the convex conjugate, this leads to

$$\widetilde{F}^*(p) \geq \langle p, cw_0 \rangle - \|cw_0\|_1 = c \left(\langle p, w_0 \rangle - \|w_0\|_1 \right).$$

Due to (3.33), for $c \to \infty$ we obtain $\widetilde{F}^*(p) = \infty$.

Consequently, the convex conjugate of $\widetilde{F}(w)$ is given by $\widetilde{F^*}(p) = \delta_{B(L^\infty)}(p)$. For

$F(w) = \alpha\widetilde{F}(w)$ we use Lemma 2.2 to see

$$F^*(p) = \alpha\delta_{B(L^\infty)}\left(\frac{p}{\alpha}\right).$$

□

Lemma 3.8. (Convex Conjugate for L^1 Distance)
For $F\left(K(w_1, w_2)\right) = \alpha_0 \left\|w_1 - w_2\right\|_1$ the convex conjugate is given by

$$F^*(p) = \alpha_0\delta_{B(L^\infty)}\left(\frac{p}{\alpha_0}\right).$$

Proof. Equivalent calculations as in Lemma 3.7 yield the asserted result. □

Lemma 3.9. (Convex Conjugate for Two L^1 Norms)
For $F\left(K(w_1, 1_2)\right) = \alpha_0 \left\|w_1 - w_2\right\|_1 + \alpha_1\|w_2\|_1$ the convex conjugate is given by

$$F^*\begin{pmatrix}p\\q\end{pmatrix} = \begin{pmatrix}\alpha_0\delta_{B(L^\infty)}\left(\frac{p}{\alpha_0}\right)\\ \alpha_1\delta_{B(L^\infty)}\left(\frac{q}{\alpha_1}\right)\end{pmatrix}.$$

Proof. Separately performing equivalent calculations as in Lemma 3.7 yields the requested result. □

3.5.3 Resolvent Operators

This section serves to calculate the explicit resolvent operators (Definition 2.23) for the different terms collected in Table 3.1. We describe the resolvent operators for G being an L^2 norm, an L^1 norm, and a combination of both. Additionally, we point out the resolvent operators for F being an L^2 norm, an L^1 norm, an affine linear L^1 norm, and an affine linear L^1 norm combined with an additional L^1 norm.

Lemma 3.10. (Resolvent Operator for Linearized L^2 Data Fidelity)
For $G(v) = \frac{1}{2}\left\|v \cdot \nabla u + u_t\right\|_2^2$ and $v : \mathbb{R}^2 \to \mathbb{R}^2$ the resolvent operator is given by

$$v_1 = \frac{a_3 b_1 - a_2 b_2}{a_1 a_3 - a_2^2}, \qquad v_2 = \frac{a_1 b_2 - a_2 b_1}{a_1 a_3 - a_2^2},$$

with

$$a_1 = 1 + \tau u_x u_x, \qquad a_2 = \tau u_x u_y, \qquad a_3 = 1 + \tau u_y u_y,$$
$$b_1 = \tilde{v_1} - \tau u_x u_t, \qquad b_2 = \tilde{v_2} - \tau u_y u_t.$$

Proof. Inserting $G(v)$ to the definition of the resolvent operator yields

$$v = \arg\min_v \left\{ \frac{\|v - \tilde{v}\|^2}{2\tau} + \frac{1}{2}\|v \cdot \nabla u + u_t\|_2^2 \right\}.$$

Since this is a pure quadratic L^2 problem, we are able to simply calculate the optimality conditions for $\boldsymbol{v} = (v_1, v_2)^T$ as follows:

$$\begin{aligned} \frac{v_1 - \tilde{v}_1}{\tau} + (v_1 u_x + v_2 u_y + u_t) u_x &\stackrel{!}{=} 0, \\ \frac{v_2 - \tilde{v}_2}{\tau} + (v_1 u_x + v_2 u_y + u_t) u_y &\stackrel{!}{=} 0. \end{aligned}$$

Rewriting this system yields

$$\begin{aligned} (1 + \tau u_x u_x) v_1 + \tau u_x u_y v_2 &= \tilde{v}_1 - \tau u_x u_t, \\ \tau u_x u_y v_1 + (1 + \tau u_y u_y) v_2 &= \tilde{v}_2 - \tau u_y u_t. \end{aligned}$$

Since all variables except for v_1 and v_2 are known in this scheme, we have a system of two equations and two unknowns. Directly solving this system with respect to v_1 and v_2 achieves the required solution. □

Lemma 3.11. (Resolvent Operator for Nonlinear L^2 Data Fidelity)
For $G(v) = \|u_0(x + \hat{v}) - u_1(x) + (v - \hat{v}) \cdot \nabla u(x + \hat{v})\|_2^2$ and $v : \mathbb{R}^2 \to \mathbb{R}^2$, the resolvent operator is given by

$$v_1 = \frac{a_3 b_1 - a_2 b_2}{a_1 a_3 - a_2^2}, \qquad v_2 = \frac{a_1 b_2 - a_2 b_1}{a_1 a_3 - a_2^2},$$

with

$$\begin{aligned} a_1 &= 1 + \tau u_x(x + \hat{v}) u_x(x + \hat{v}), \\ a_2 &= \tau u_x(x + \hat{v}) u_y(x + \hat{v}), \\ a_3 &= 1 + \tau u_y(x + \hat{v}) u_y(x + \hat{v}), \\ b_1 &= \tilde{v_1} - \tau \left(-\hat{v} u_x(x + \hat{v}) - \hat{v} u_y(x + \hat{v}) + u_0(x + \hat{v}) - u_1(x) \right) u_x(x + \hat{v}), \\ b_2 &= \tilde{v_2} - \tau \left(-\hat{v} u_x(x + \hat{v}) - \hat{v} u_y(x + \hat{v}) + u_0(x + \hat{v}) - u_1(x) \right) u_x(x + \hat{v}). \end{aligned}$$

Proof. The calculations are equivalent to Lemma 3.10. □

Lemma 3.12. (Resolvent Operator for Linearized L^1 Data Fidelity)
For $G(v) = \|v \cdot \nabla u + u_t\|_1$ the resolvent operator is given by

$$v = \tilde{v} + \begin{cases} -\tau\nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u)^2 \\ \tau\nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u)^2 \\ -\frac{\rho(\tilde{v})}{\nabla u} & \text{if } |\rho(\tilde{v})| \leq \tau(\nabla u)^2 \end{cases},$$

with $\rho(v) := v \cdot \nabla u + u_t$.

Proof. Inserting $G(v)$ to the definition of the resolvent operator yields

$$v = \arg\min_v \left\{ \frac{\|v - \tilde{v}\|^2}{2\tau} + \|v \cdot \nabla u + u_t\|_1 \right\}.$$

Since the L^1 norm is not Fréchet differentiable in $v \cdot \nabla u + u_t = 0$, we have to distinguish between the cases $v < -\frac{u_t}{\nabla u}$, $v > -\frac{u_t}{\nabla u}$ and $v = -\frac{u_t}{\nabla u}$.

First case: $v < -\frac{u_t}{\nabla u}$
Since the inside of the L^1 norm is completely negative in this case, we are able to calculate the optimality condition as follows:

$$\frac{v - \tilde{v}}{\tau} - \nabla u \stackrel{!}{=} 0.$$

Thus, the solution for v can be written as

$$v = \tilde{v} + \tau\nabla u.$$

To generalize the distinction of cases with respect to the optical flow constraint, we use this solution and insert it into the condition $v < -\frac{u_t}{\nabla u}$:

$$\tilde{v} + \tau\nabla u < -\frac{u_t}{\nabla u} \qquad \Leftrightarrow \qquad \tilde{v} \cdot \nabla u + u_t < -\tau(\nabla u)^2.$$

Consequently, for $\rho(v) = v \cdot \nabla u + u_t$ this first case is equivalent to

$$\rho(\tilde{v}) < -\tau(\nabla u)^2.$$

Second case: $v > -\frac{u_t}{\nabla u}$
This case can be treated analogously to the first case.
Third case: $v = -\frac{u_t}{\nabla u}$
Since there is no Fréchet derivative in this case, we need an element d of the subdifferential

of $\|v \cdot \nabla u + u_t\|_1$ to calculate the optimality condition. Due to the multiplication of v with ∇u, the subdifferential is given by $\{p \in [-\nabla u, \nabla u]\}$. With such p, the optimality condition is

$$\frac{v - \tilde{v}}{\tau} + p \stackrel{!}{=} 0 \qquad \Leftrightarrow \qquad v = \tilde{v} - \tau p.$$

To compute the exact p, we use the condition $v = -\frac{u_t}{\nabla u}$ and insert it into the optimality condition:

$$\frac{-\frac{u_t}{\nabla u} - \tilde{v}}{\tau} + p = 0 \qquad \Leftrightarrow \qquad p = \frac{u_t}{\tau \nabla u} + \frac{\tilde{v}}{\tau}.$$

With this p the solution for v reads

$$v = \tilde{v} - \tau \left(\frac{u_t}{\tau \nabla u} + \frac{\tilde{v}}{\tau} \right) = \tilde{v} - \left(\frac{u_t}{\nabla u} + \tilde{v} \right) = \tilde{v} - \frac{\rho(\tilde{v})}{\nabla u}.$$

Again, to formulate a condition for $\rho(\tilde{v})$ as defined in the first case, we use the fact that $|p| \leq \nabla u$ to identify

$$\left| \frac{u_t}{\tau \nabla u} + \frac{\tilde{v}}{\tau} \right| \leq \nabla u \qquad \Leftrightarrow \qquad |\tilde{v} \cdot \nabla u + u_t| \leq \tau (\nabla u)^2 \qquad \Leftrightarrow \qquad |\rho(\tilde{v})| \leq \tau (\nabla u)^2.$$

Combining all cases, we achieve the requested solution for v. □

Lemma 3.13. (Resolvent Operator for Nonlinear L^2 Data Fidelity)
For $G(v) = \|u_0(x + \hat{v}) - u_1(x) + (v - \hat{v}) \cdot \nabla u(x + \hat{v})\|_1$, the resolvent operator is given by

$$v = \tilde{v} + \begin{cases} -\tau \nabla u & \text{if } \rho(\tilde{v}) < -\tau (\nabla u\,(x + \hat{v}))^2 \\ \tau \nabla u & \text{if } \rho(\tilde{v}) < \tau (\nabla u\,(x + \hat{v}))^2 \\ -\frac{\rho(\tilde{v})}{\nabla u(x + \tilde{v})} & \text{if } |\rho(\tilde{v}\,)| \leq \tau (\nabla u(x + \hat{v}))^2 \end{cases},$$

with $\rho(v) \coloneqq u_0(x + \hat{v}) - u_1(x) + (v - \tilde{v}) \cdot \nabla u(x + \tilde{v})$.

Proof. The calculations are equivalent to those in Lemma 3.12. □

For the resolvent operators that contain v and w we split the input $\tilde{v}$ into those parts that are used for v and those that are used for w, and continue denoting these variables with $\tilde{v}$ and $\tilde{w}$.

Lemma 3.14. (Resolvent Operator for Linearized L^2 Data Fidelity with Extention)
For $G(v, w) = \frac{1}{2} \|v \cdot \nabla u + u_t\|_2^2 + \frac{\alpha_1}{2} \|w\|_2^2$ and $v : \mathbb{R}^2 \to \mathbb{R}^2$ the resolvent operators are

given by

$$v_1 = \frac{(\tilde{v}_1 - \tau u_x u_t)(1 + \tau u_y u_y) - (\bar{v}_2 - \tau u_y u_t)(\tau u_x u_y)}{(1 + \tau u_x u_x)(1 + \tau u_y u_y) - (\tau u_x u_y)^2},$$
$$v_2 = \frac{(\tilde{v}_2 - \tau u_y u_t)(1 + \tau u_x u_x) - (\bar{v}_1 - \tau u_x u_t)(\tau u_x u_y)}{(1 + \tau u_x u_x)(1 + \tau u_y u_y) - (\tau u_x u_y)^2},$$

and

$$w = \frac{\tilde{w}}{1 + \tau\alpha_1}.$$

Proof. Inserting $G(v, w)$ to the definition of the resolvent operator yields

$$\begin{aligned} v &= \arg\min_v \left\{ \frac{\|v - \tilde{v}\|^2}{2\tau} + \frac{1}{2}\|v \cdot \nabla u + u_t\|_2^2 + \frac{\alpha_1}{2}\|w\|_2^2 \right\} \\ &= \arg\min_v \left\{ \frac{\|v - \tilde{v}\|^2}{2\tau} + \frac{1}{2}\|v \cdot \nabla u + u_t\|_2^2 \right\}, \end{aligned}$$

and

$$\begin{aligned} w &= \arg\min_w \left\{ \frac{\|w - \tilde{w}\|^2}{2\tau} + \frac{1}{2}\|v \cdot \nabla u + u_t\|_2^2 + \frac{\alpha_1}{2}\|w\|_2^2 \right\} \\ &= \arg\min_w \left\{ \frac{\|w - \tilde{w}\|^2}{2\tau} + \frac{\alpha_1}{2}\|w\|_2^2 \right\}. \end{aligned}$$

The equation for v is equivalent to the result in Lemma 3.10 and thus we end up in

$$v_1 = \frac{(\tilde{v}_1 - \tau u_x u_t)(1 + \tau u_y u_y) - (\tilde{v}_2 - \tau u_y u_t)(\tau u_x u_y)}{(1 + \tau u_x u_x)(1 + \tau u_y u_y) - (\tau u_x u_y)^2},$$
$$v_2 = \frac{(\tilde{v}_2 - \tau u_y u_t)(1 + \tau u_x u_x) - (\tilde{v}_1 - \tau u_x u_t)(\tau u_x u_y)}{(1 + \tau u_x u_x)(1 + \tau u_y u_y) - (\tau u_x u_y)^2}.$$

Calculating the optimality condition for the equation for w yields

$$\frac{w - \tilde{w}}{\tau} + \alpha_1 w \stackrel{!}{=} 0 \qquad \Leftrightarrow \qquad w = \frac{\tilde{w}}{1 + \tau\alpha_1}.$$

□

Lemma 3.15. (Resolvent Operator for Nonlinear L^2 Data Fidelity with Extention) For $G(v) = \|u_0(x + \hat{v}) - u_1(x) + (v - \hat{v}) \cdot \nabla u(x + \hat{v})\|_2^2 + \frac{\alpha_1}{2}\|w\|_2^2$ and $v : \mathbb{R}^2 \to \mathbb{R}^2$ the resolvent operators are given by

$$v_1 = \frac{a_3 b_1 - a_2 b_2}{a_1 a_3 - a_2^2}, \qquad v_2 = \frac{a_1 b_2 - a_2 b_1}{a_1 a_3 - a_2^2},$$

with

$$
\begin{aligned}
a_1 &= 1 + \tau u_x(x+\hat{v})u_x(x+\hat{v}),\\
a_2 &= \tau u_x(x+\hat{v})u_y(x+\hat{v}),\\
a_3 &= 1 + \tau u_y(x+\hat{v})u_y(x+\hat{v}),\\
b_1 &= \tilde{v}_1 - \tau\left(-\hat{v}u_x(x+\hat{v}) - \hat{v}u_y(x+\hat{v}) + u_0(x+\hat{v}) - u_1(x)\right)u_x(x+\hat{v}),\\
b_2 &= \tilde{v}_2 - \tau\left(-\hat{v}u_x(x+\hat{v}) - \hat{v}u_y(x+\hat{v}) + u_0(x+\hat{v}) - u_1(x)\right)u_x(x+\hat{v}),
\end{aligned}
$$

and

$$
w = \frac{\tilde{w}}{1+\tau\alpha_1}.
$$

Proof. The justification is equivalent to Lemma 3.14. □

Lemma 3.16. (Resolvent Operator for Linearized L^1 Data Fidelity with Extention) For $G(v,w) = \|v\cdot\nabla u + u_t\|_1 + \frac{\alpha_1}{2}\|w\|_2^2$ the resolvent operators are given by

$$
v = \tilde{v} + \begin{cases} -\tau\nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u)^2 \\ \tau\nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u)^2 \\ -\frac{\rho(\tilde{v})}{\nabla u} & \text{if } |\rho(\tilde{v})| \le \tau(\nabla u)^2 \end{cases},
$$

with $\rho(v) \coloneqq v\cdot\nabla u + u_t$, and

$$
w = \frac{\tilde{w}}{1+\tau\alpha_1}.
$$

Proof. Inserting $G(v,w)$ to the definition of the resolvent operator yields

$$
\begin{aligned}
v &= \arg\min_v \left\{ \frac{\|v-\tilde{v}\|^2}{2\tau} + \|v\cdot\nabla u + u_t\|_1 + \frac{\alpha_1}{2}\|w\|_2^2 \right\}\\
&= \arg\min_v \left\{ \frac{\|v-\tilde{v}\|^2}{2\tau} + \|v\cdot\nabla u + u_t\|_1 \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
w &= \arg\min_w \left\{ \frac{\|w-\tilde{w}\|^2}{2\tau} + \|v\cdot\nabla u + u_t\|_1 + \frac{\alpha_1}{2}\|w\|_2^2 \right\}\\
&= \arg\min_w \left\{ \frac{\|w-\tilde{w}\|^2}{2\tau} + \frac{\alpha_1}{2}\|w\|_2^2 \right\}.
\end{aligned}
$$

The equation for v is equivalent to the result in Lemma 3.12 and thus we end up in

$$v = \tilde{v} + \begin{cases} -\tau\nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u)^2 \\ \tau\nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u)^2 \\ -\frac{\rho(\tilde{v})}{\nabla u} & \text{if } |\rho(\tilde{v})| \leq \tau(\nabla u)^2 \end{cases},$$

with $\rho(v) = v \cdot \nabla u + u_t$. The equation for w is equivalent to the subproblem in w in Lemma 3.14. Thus, we obtain

$$w = \frac{\tilde{w}}{1 + \tau\alpha_1}.$$

□

Lemma 3.17. (Resolvent Operator for Nonlinear L^1 Data Fidelity with Extention)
For $G(v, w) = \|u_0(x + \hat{v}) - u_1(x) + (v - \hat{v}) \cdot \nabla u(x + \hat{v})\|_1 + \frac{\alpha_1}{2}\|w\|_2^2$ the resolvent operators are given by

$$v = \tilde{v} + \begin{cases} -\tau\nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u\,(x + \hat{v}))^2 \\ \tau\nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u\,(x + \hat{v}))^2 \\ -\frac{\rho(\tilde{v})}{\nabla u(x+\tilde{v})} & \text{if } |\rho(\tilde{v}\,)| \leq \tau(\nabla u(x + \hat{v}))^2 \end{cases},$$

with $\rho(v) \coloneqq u_0(x + \hat{v}) - u_1(x) + (v - \tilde{v}) \cdot \nabla u(x + \tilde{v})$, and

$$w = \frac{\hat{w}}{1 + \tau\alpha_1}.$$

Proof. The justification is equivalent to Lemma 3.16. □

Table 3.2 summarizes the results for the resolvent operators of the single terms that occur in $G(v)$ and $G(v, w)$.

Lemma 3.18. (Resolvent Operator for Squared L^2 Norm)
For $F(Kv) = \frac{\alpha}{2}\|\nabla v\|_2^2$ the adjoint resolvent operator is given by

$$p = \frac{\alpha}{\alpha + \sigma}\tilde{p},$$

where p is the dual variable of v.

Proof. From Lemma 3.6 we know the convex conjugate of the given functional:

$$F^*(p) = \frac{1}{2\alpha}\|p\|_2^2.$$

G	v/w	
$\mathcal{D}_{Lin2}(u, \boldsymbol{v})$	$\begin{pmatrix} \frac{a_3 b_1 - a_2 b_2}{a_1 a_3 - a_2^2} \\ \frac{a_1 b_2 - a_2 b_1}{a_1 a_3 - a_2^2} \end{pmatrix}$	$a_1 = 1 + \tau u_x u_x$ $a_2 = \tau u_x u_y$ $a_3 = 1 + \tau u_y u_y$ $b_1 = \tilde{v_1} - \tau u_x u_t$ $b_2 = \tilde{v_2} - \tau u_y u_t$
$\mathcal{D}_{NL2}(u, \boldsymbol{v})$	$\begin{pmatrix} \frac{a_3 b_1 - a_2 b_2}{a_1 a_3 - a_2^2} \\ \frac{a_1 b_2 - a_2 b_1}{a_1 a_3 - a_2^2} \end{pmatrix}$	$a_1 = 1 + \tau u_x(x+\hat{v}) u_x(x+\hat{v})$ $a_2 = \tau u_x(x+\hat{v}) u_y(x+\hat{v})$ $a_3 = 1 + \tau u_y(x+\hat{v}) u_y(x+\hat{v})$ $b_1 = \tilde{v_1} - \tau\left(-\hat{v} u_x(x+\hat{v}) - \hat{v} u_y(x+\hat{v})\right.$ $\left. + u_0(x+\hat{v}) - u_1(x)\right) u_x(x+\hat{v})$ $b_2 = \tilde{v_2} - \tau\left(-\hat{v} u_x(x+\hat{v}) - \hat{v} u_y(x+\hat{v})\right.$ $\left. + u_0(x+\hat{v}) - u_1(x)\right) u_x(x+\hat{v})$
$\mathcal{D}_{Lin1}(u, \boldsymbol{v})$	$\tilde{v} + \begin{cases} -\tau \nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u)^2 \\ \tau \nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u)^2 \\ -\frac{\rho(\tilde{v})}{\nabla u} & \text{if } \lvert\rho(\tilde{v})\rvert \leq \tau(\nabla u)^2 \end{cases}$ $\rho(v) = v \cdot \nabla u + u_t$	
$\mathcal{D}_{NL1}(u, \boldsymbol{v})$	$\tilde{v} + \begin{cases} -\tau \nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u\,(x+\hat{v}))^2 \\ \tau \nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u\,(x+\hat{v}))^2 \\ -\frac{\rho(\tilde{v})}{\nabla u(x+\tilde{v})} & \text{if } \lvert\rho(\tilde{v}\,)\rvert \leq \tau(\nabla u(x+\hat{v}))^2 \end{cases}$ $\rho(v) = u_0(x+\hat{v}) - u_1(x) + (v-\tilde{v}) \cdot \nabla u(x+\tilde{v})$	
$\frac{\alpha_1}{2}\lVert w\rVert_2^2$	$\frac{\tilde{w}}{1+\tau\alpha_1}$	
$\frac{\alpha_1}{2}\lVert w\rVert_2^2$	$\frac{\tilde{w}}{1+\tau\alpha_1}$	

Table 3.2: Resolvent operators for the single terms of G.

Inserting this into the definition of the resolvent operator yields

$$p = \arg\min_{p} \left\{ \frac{\|p - \tilde{p}\|^2}{2\sigma} + \frac{1}{2\alpha}\|p\|_2^2 \right\}.$$

Equivalent to the subproblem in w in Lemma 3.14, we calculate the optimality condition and obtain

$$p = \frac{\alpha}{\alpha + \sigma}\tilde{p}.$$

□

Lemma 3.19. (Resolvent Operator for L^1 Norm)
For $F(Kv) = \alpha \|\nabla v\|_1$ the adjoint resolvent operator is given by

$$p = \pi_\alpha(\tilde{p}) := \min_{\tilde{p}} \left(\alpha, \max_{\tilde{p}}(-\alpha, \tilde{p}) \right).$$

where p is the dual variable of v.

Proof. We obtain the convex conjugate of the given functional from Lemma 3.7:

$$F^*(p) = \alpha \delta_{B(L^\infty)} \left(\frac{p}{\alpha} \right).$$

Inserting this into the definition of the resolvent operator yields

$$p = \arg\min_{p} \left\{ \frac{\|p - \tilde{p}\|^2}{2\sigma} + \alpha \delta_{B(L^\infty)} \left(\frac{p}{\alpha} \right) \right\}. \tag{3.34}$$

To define the solution to this optimality problem, we need to identify those elements $\left(\frac{p}{\alpha}\right)$, which are in the L^∞ unit ball, i.e.

$$\left\| \frac{p}{\alpha} \right\|_\infty \leq 1 \qquad \Leftrightarrow \qquad \|p\|_\infty \leq \alpha.$$

With the definition of the norm in L^∞ it follows

$$\alpha \geq \|p\|_\infty = \max_{i,j} |p_{i,j}|.$$

If this condition is fulfilled, the second term in (3.34) is equal to zero. The first term is zero if $p = \tilde{p}$. Due to the positivity of the single terms we find a minimum of the problem if p is chosen as close to $\tilde{p}$ as possible, but still within $[-\alpha, \alpha]$. With

$$\pi_\alpha(\tilde{p}) := \min_{\tilde{p}} \left(\alpha, \max_{\tilde{p}}(-\alpha, \tilde{p}) \right)$$

$\mathbf{F}$	$\mathbf{p}$	$\mathbf{q}$	$\mathbf{K}^*$
$\frac{\alpha}{2}\left\|\nabla v\right\|_2^2$	$\frac{\alpha}{\alpha+\sigma}\tilde{p}$		$-\nabla\cdot$
$\alpha\left\|\nabla v\right\|_1$	$\pi_\alpha(\tilde{p})$		$-\nabla\cdot$
$\alpha_0\left\|\nabla v - w\right\|_1$	$\pi_{\alpha_0}(\tilde{p})$		$-\begin{pmatrix}\nabla\cdot\\ I\end{pmatrix}$
$\alpha_0\left\|\nabla v - w\right\|_1 + \alpha_1\|\nabla w\|_1$	$\pi_{\alpha_0}(\tilde{p})$	$\pi_{\alpha_1}(\tilde{q})$	$-\begin{pmatrix}\nabla\cdot & 0\\ I & \nabla\cdot\end{pmatrix}$

Table 3.3: Resolvent and adjoint operators for F.

we define a point-wise projection of $\tilde{p}$ onto $[-\alpha, \alpha]$, and see that the required minimum is given by $p = \pi_\alpha(\tilde{p})$. □

Lemma 3.20. (Resolvent Operator for Affine Linear L^1 Norm)
For $F\left(K(v,w)\right) = \alpha_0\left\|\nabla v - w\right\|_1$ the adjoint resolvent operator is given by

$$p = \pi_{\alpha_0}(\tilde{p}) \coloneqq \min_{\tilde{p}}\left(\alpha_0, \max_{\tilde{p}}(-\alpha_0, \tilde{p})\right),$$

where p is the dual variable of $(v,w)^T$.

Proof. Using the convex conjugate from Lemma 3.8 and performing equivalent calculations as in Lemma 3.19 leads to the requested result. □

Lemma 3.21. (Resolvent Operator for Affine Linear L^1 Norm with Extension)
For $F\left(K(v,w)\right) = \alpha_0\left\|\nabla v - w\right\|_1 + \alpha_1\|w\|_1$ the adjoint resolvent operators are given by

$$p = \pi_{\alpha_0}(\tilde{p}) \coloneqq \min_{\tilde{p}}\left(\alpha_0, \max_{\tilde{p}}(-\alpha_0, \tilde{p})\right),$$
$$q = \pi_{\alpha_1}(\tilde{q}) \coloneqq \min_{\tilde{q}}\left(\alpha_1, \max_{\tilde{q}}(-\alpha_1, \tilde{q})\right)$$

where p is the dual variable of v and q is the dual variable of w.

Proof. Using the convex conjugate from Lemma 3.9 and performing equivalent calculations as in Lemma 3.19 leads to the requested result. □

Table 3.3 summarizes the results for the resolvent operators for the different choices of F and the particular adjoint operators of K.

3.6 Discretization

For the sake of simplicity, we only show the discretization for two-dimensional images in this chapter. However, the discretization for three dimensions is straight forward and can be calculated equivalently. For the temporal discretization we denote the time steps by $0, \cdots, T$ with step length δ_t. For the spatial discretization we define equidistant grids with grid points $0, \cdots, n_x$ with step length δ_x in x-direction, and grid points $0, \cdots, n_y$ with step length δ_y in y-direction. The displacement of a certain pixel (i, j) at a distinct time step, is described by (Δ_i, Δ_j). Since we need a temporal derivative for u in the data fidelity, the discrete Version U of u is given given at

$$\left\{(i,j,t)\middle| i = 0, \cdots, n_x, j = 0, \cdots, n_y, t = 0, \cdots, T\right\}.$$

The remaining discrete variables V of v, W of w, P of p, and Q of q do not require a temporal grid and are thus given at

$$\left\{(i,j)\middle| i = 0, \cdots, n_x, j = 0, \cdots, n_y\right\}.$$

Nevertheless, these points also need to be evaluated at every time step.

For the spatial and temporal derivatives of U, V, W, P and Q, we stick to the discretizations proposed in Dirks (2015). This particular combination of discretizations for the different variables turns out to be stable for variables that have equivalent characteristics as those used in this chapter. For the temporal derivative U_t, we use forward differences with Dirichlet boundary conditions, i.e.

$$U_t(i,j,t) = \frac{U(i,j,t+1) - U(i,j,t)}{\delta_t}.$$

For the Operator ∇ we distinguish between an application to the image U, and an application to the velocity field V or W. The reason for this is the different characteristic of the equation depending on the viewpoint of the variables, see Dirks (2015). For ∇U we use central differences, i.e.

$$U_x(i,j,t) = \begin{cases} \frac{U(i+1,j,t) - U(i-1,j,t)}{2\delta_x} & \text{if } i > 0 \text{ and } i < n_x \\ 0 & \text{if } i = 0 \text{ or } i = n_x \end{cases},$$

$$U_y(i,j,t) = \begin{cases} \frac{U(i,j+1,t) - U(i,j-1,t)}{2\delta_y} & \text{if } j > 0 \text{ and } j < n_y \\ 0 & \text{if } j = 0 \text{ or } j = n_y \end{cases}.$$

For ∇V and ∇W we use forward differences with Dirichlet boundary conditions:

$$V_x(i,j) = \begin{cases} \frac{V(i+1,j)-V(i,j)}{\delta_x} & \text{if } i < n_x \\ 0 & \text{if } i = n_x \end{cases},$$

$$V_y(i,j) = \begin{cases} \frac{V(i,j+1)-V(i,j)}{\delta_y} & \text{if } j < n_y \\ 0 & \text{if } j = n_y \end{cases},$$

$$W_x(i,j) = \begin{cases} \frac{W(i+1,j)-W(i,j)}{\delta_x} & \text{if } i < n_x \\ 0 & \text{if } i = n_x \end{cases},$$

$$W_y(i,j) = \begin{cases} \frac{W(i,j+1)-W(i,j)}{\delta_y} & \text{if } j < n_y \\ 0 & \text{if } j = n_y \end{cases}.$$

Consequently, for the adjoint operators $\nabla \cdot P$ and $\nabla \cdot Q$ we use backwards differences with Neumann boundary conditions:

$$P_x(i,j) = \begin{cases} \frac{P(i,j)-P(i-1,j)}{\delta_x} & \text{if } i > 0 \text{ and } i < n_x \\ P(i,j) & \text{if } i = 0 \\ -P(i-1,j) & \text{if } i = n_x \end{cases},$$

$$P_y(i,j) = \begin{cases} \frac{P(i,j)-P(i,j-1)}{\delta_y} & \text{if } j > 0 \text{ and } j < n_y \\ P(i,j) & \text{if } j = 0 \\ -P(i,j-1) & \text{if } j = n_y \end{cases},$$

$$Q_x(i,j) = \begin{cases} \frac{Q(i,j)-Q(i-1,j)}{\delta_x} & \text{if } i > 0 \text{ and } i < n_x \\ Q(i,j) & \text{if } i = 0 \\ -Q(i-1,j) & \text{if } i = n_x \end{cases},$$

$$Q_y(i,j) = \begin{cases} \frac{Q(i,j)-Q(i,j-1)}{\delta_y} & \text{if } j > 0 \text{ and } j < n_y \\ Q(i,j) & \text{if } j = 0 \\ -Q(i,j-1) & \text{if } j = n_y \end{cases}.$$

With the previously described definitions, the discretized version of the brightness constancy assumption reads as follows:

$$U(i+\Delta_i, j+\Delta_j, t) = U(i,j,t+\delta_t). \tag{3.35}$$

In the discrete setting the linearized optical flow constraint can be derived by an approximation of $U(i+\Delta_i, j+\Delta_j, t)$ via Taylor expansion at $(x, y, t+\delta_t)$:

$$\begin{aligned} U(i+\Delta_i, j+\Delta_j, t) =& U(i,j,t+\delta_t) + U_x(i,j,t+\delta_t)\Delta i + U_y(i,j,t+\delta_t)\Delta j \\ &+ U_t(i,j,t+\delta_t)(-\delta_t) + \text{h.o.t. }. \end{aligned}$$

In case that the displacements Δ_i and Δ_j are small enough, i.e. smaller than the grid size, the higher order terms are small enough such that they can be neglected. We insert the remaining terms of the above equation into (3.35) and obtain

$$\begin{aligned} U(i,j,t+\delta_t) =& U(i,j,t+\delta_t) + U_x(i,j,t+\delta_t)\Delta_i + U_y(i,j,t+\delta_t)\Delta_j \\ &+ U_t(i,j,t+\delta_t)(-\delta_t), \end{aligned}$$

which is equivalent to

$$0 = U_x(i,j,t+\delta_t)\Delta_i + U_y(i,j,t+\delta_t)\Delta_j + U_t(i,j,t+\delta_t)(-\delta_t).$$

The velocity field $V(i,j)$ at $t+\delta_t$ is given by $V(i,j) = -\frac{(\Delta_i, \Delta_j)}{\delta_t}$. Hence, associating U_x and U_y with ∇U, the discrete version of the linearized optical flow constraint is given by

$$0 = U_t(i,j,t+\delta_t) + \nabla U(i,j,t+\delta_t) \cdot V(i,j).$$

The nonlinear version of the brightness constancy assumption can be identified in the discrete setting by

$$U(V(i,j),t) = U(i,j,t+\delta_t).$$

The Gauß-Newton method (3.4), which we introduced in Section 3.2.2 for solving the nonlinear optical flow constraint, can also be derived in the discrete setting. Therefore, we start with the linearized brightness constancy assumption (3.35). We introduce an approximation $\left(\widehat{\Delta}_i, \widehat{\Delta}_j\right)$ of the displacement (Δ_i, Δ_j), i.e.

$$\left| \left(\widehat{\Delta}_i, \widehat{\Delta}_j\right) - (\Delta_i, \Delta_j) \right| < (\delta_x, \delta_y). \tag{3.36}$$

The idea for this adapted optical flow constraint is based on Zach et al. (2007). The approximation of the displacement can be achieved iteratively by performing a multiscale approach. Similar to the linearized optical flow constraint, we approximate

$U(i+\Delta_i, j+\Delta_j, t)$ with a Taylor expansion. However, to avoid large higher order terms, which appear in case that the covered distances (Δ_i, Δ_j) are larger than the grid size (δ_i, δ_j), we evaluate the expansion at $\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big)$ instead of $(i, j, t+\delta_t)$:

$$\begin{aligned}
U(i+\Delta_i, j+\Delta_j, t) =& U\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) \\
&+ U_x\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) \cdot \Big((i+\Delta_i) - \big(i+\widehat{\Delta}_i\big)\Big) \\
&+ U_y\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) \cdot \Big((j+\Delta_j) - \big(j+\widehat{\Delta}_j\big)\Big) \\
&+ U_t\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) \cdot (t-t) + \text{h.o.t.} \\
=& U\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t+\delta_t\big) \\
&+ U_x\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t+\delta_t\big) \cdot \big(\Delta_i - \widehat{\Delta}_i\big) \\
&+ U_y\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t+\delta_t\big) \cdot \big(\Delta_j - \widehat{\Delta}_j\big) + \text{h.o.t.}\ .
\end{aligned}$$

Due to (3.36), the higher order terms are small and can be ignored even if the actual distances (Δ_i, Δ_j) can be large. We insert the remaining terms into (3.35) and obtain:

$$\begin{aligned}
U(i, j, t+\delta_t) =& U\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) + U_x\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) \cdot \big(\Delta_i - \widehat{\Delta}_i\big) \\
&+ U_y\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) \cdot \big(\Delta_j - \widehat{\Delta}_j\big).
\end{aligned}$$

Again, associating U_x and U_y with ∇U gives us:

$$0 = U\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big) - U(i, j, t+\delta_t) + \Big((\Delta_i, \Delta_j) - \big(\widehat{\Delta}_i, \widehat{\Delta}_j\big)\Big) \nabla U\big(i+\widehat{\Delta}_i, j+\widehat{\Delta}_j, t\big).$$

Identifying the actual velocity field by $V(i, j) \coloneqq (\Delta_i, \Delta_j)$, and the approximated velocity field by $\hat{V}(i, j) \coloneqq \big(\widehat{\Delta}_i, \widehat{\Delta}_j\big)$, we obtain the discrete Gauß-Newton method:

$$\begin{aligned}
0 =& U\Big((i, j) + \hat{V}(i, j), t\Big) - U\left((i, j), t+\delta_t\right) \\
&+ \Big(V(i, j) - \hat{V}(i, j)\Big) \nabla U\Big((i, j) + \hat{V}(i, j), t\Big).
\end{aligned}$$

3.7 Numerical Realization

The code we use for the implementation of the optical flow models is based the Motion Estimation GUI by Hendrik Dirks, which can be downloaded at Dirks (2016a). However, we adapted the code with respect to our requirements, e.g. for including the nonlinear data fidelities and the Bregman iterations.

The performance of the models including a nonlinear optical flow constraint is dependent on several additional parameters compared to the linearized models. This is a

consequence of the warping step, which serves to approximate the velocity field at the current grid size. To nevertheless enable an equal comparison between the different fidelities, we forgo a parameter optimization for these additional possibilities. We rather stick to a fixed adjustment for all of the following approaches. The specific choices for all parameters are equal to those in Dirks (2016b). Naturally, the occurring regularization parameters are excluded from this fixation.

3.7.1 Iteration Scheme

For the numerical implementation, we use the Chambolle-Pock algorithm introduced in Definition 2.24. The original algorithm only considers a single primal and a single dual variable. Therefore, we need to adapt it in order to include the additional variables that we introduced for the extended regularizers in Section 3.3.2.

Definition 3.3. (Adapted Chambolle-Pock Algorithm)
Choose $\tau, \sigma > 0$, $\theta \in [0,1]$, $(u^0, w^0, p^0, q^0) \in \mathcal{X} \times \mathcal{Y}$ and set $\bar{u}^0 = u^0$ and $\bar{w}^0 = w^0$. For $n \geq 0$ the updates of $u^n, w^n, p^n, q^n, \bar{u^n}$ and $\bar{w^n}$ are defined as follows:

$$\begin{aligned}
\begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix} &= (I + \sigma \partial F^*)^{-1} \left(\begin{pmatrix} p^n \\ q^n \end{pmatrix} + \sigma K \begin{pmatrix} \bar{u}^n \\ \bar{w}^n \end{pmatrix} \right), \\
\begin{pmatrix} u^{n+1} \\ w^{n+1} \end{pmatrix} &= (I + \tau \partial G)^{-1} \left(\begin{pmatrix} u^n \\ w^n \end{pmatrix} - \tau K^* \begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix} \right), \\
\begin{pmatrix} \bar{u}^{n+1} \\ \bar{w}^{n+1} \end{pmatrix} &= \begin{pmatrix} u^{n+1} \\ w^{n+1} \end{pmatrix} + \theta \left(\begin{pmatrix} u^{n+1} \\ w^{n+1} \end{pmatrix} - \begin{pmatrix} u^n \\ w^n \end{pmatrix} \right).
\end{aligned}$$

In order to be able to apply the resolvent operators $(I + \sigma \partial F^*)^{-1}$ and $(I + \tau \partial G)^{-1}$ to the particular input, we calculate them in each iteration separately. Thus, for $n > 1$ we

obtain the following scheme:

$$\begin{pmatrix} \tilde{p}^{n+1} \\ \tilde{q}^{n+1} \end{pmatrix} = \begin{pmatrix} p^n \\ q^n \end{pmatrix} + \sigma K \begin{pmatrix} \bar{u}^n \\ \bar{w}^n \end{pmatrix}, \tag{3.37}$$

$$p^{n+1} = \arg\min_p \left\{ \frac{\|p - \tilde{p}^{n+1}\|^2}{2\sigma} + F^*(p) \right\}, \tag{3.38}$$

$$q^{n+1} = \arg\min_q \left\{ \frac{\|q - \tilde{q}^{n+1}\|^2}{2\sigma} + F^*(q) \right\}, \tag{3.39}$$

$$\begin{pmatrix} \tilde{u}^{n+1} \\ \tilde{w}^{n+1} \end{pmatrix} = \begin{pmatrix} u^n \\ w^n \end{pmatrix} - \tau K^* \begin{pmatrix} p^{n+1} \\ q^{n+1} \end{pmatrix}, \tag{3.40}$$

$$u^{n+1} = \arg\min_u \left\{ \frac{\|u - \tilde{u}^{n+1}\|^2}{2\tau} + G(u) \right\}, \tag{3.41}$$

$$w^{n+1} = \arg\min_w \left\{ \frac{\|w - \tilde{w}^{n+1}\|^2}{2\tau} + G(w) \right\}, \tag{3.42}$$

$$\begin{pmatrix} \bar{u}^{n+1} \\ \bar{w}^{n+1} \end{pmatrix} = \begin{pmatrix} u^{n+1} \\ w^{n+1} \end{pmatrix} + \theta \left(\begin{pmatrix} u^{n+1} \\ w^{n+1} \end{pmatrix} - \begin{pmatrix} u^n \\ w^n \end{pmatrix} \right). \tag{3.43}$$

We use this scheme for all of the analyzed models. For those models that do not contain the variables w or q, we simply ignore (3.39) or (3.42) and the respective parts of the vectors in (3.37), (3.40), and (3.43). The operator K in (3.37) is defined in Table 3.1 for the single models. The corresponding adjoint operator K^* in (3.40) is given in Section 3.5.1. For the resolvent operators in (3.38), (3.39), (3.41), and (3.42), we replace the equations by the particular results of Section 3.5.3.

3.7.2 Initialization and Stopping Criterion

For the initialization we set

$$v = \bar{v} = w = \bar{w} = p = q = 0.$$

Additionally, we set $\Theta = 1$ and $\tau = \sigma = \frac{1}{\sqrt{8}}$. In Chambolle (2004) it was shown that this choice for τ and σ fulfills the condition for convergence, which we mentioned in Remark 2.12.

For the stopping criterion we use the primal-dual residual proposed in Goldstein et al. (2013). For models with standard regularizers the primal residual $\tilde{P}$ and dual residual

$\widetilde{D}$ are given by

$$\widetilde{P}_{n+1}(v,p) \coloneqq \frac{1}{\tau_n}(v_k - v_{n+1}) - K^*(p_n - p_{n+1}),$$
$$\widetilde{D}_{n+1}(v,p) \coloneqq \frac{1}{\sigma_n}(p_k - p_{n+1}) - K^*(v_n - v_{n+1}).$$

In Goldstein et al. (2013) it is shown that the algorithm converges, if

$$\lim_{n\to\infty} \|P_n\|^2 + \|D_n\|^2 = 0.$$

Since we want to use the same stopping criterion for both standard and extended regularizers, we need to extend the residuals and obtain:

$$P_{n+1}(v,w,p,q) =\coloneqq \frac{1}{\tau_n}\begin{pmatrix} v_n - v_{n+1} \\ w_n - w_{n+1} \end{pmatrix} - K^*\begin{pmatrix} p_n - p_{n+1} \\ q_n - q_{n+1} \end{pmatrix},$$
$$D_{n+1}(v,w,p,q) =\coloneqq \frac{1}{\sigma_n}\begin{pmatrix} p_n - p_{n+1} \\ q_n - q_{n+1} \end{pmatrix} - K^*\begin{pmatrix} v_n - v_{n+1} \\ w_n - w_{n+1} \end{pmatrix}.$$

We stop the iterations if the error

$$\epsilon_n \coloneqq \frac{|P_n| + |D_n|}{|\Omega|} \tag{3.44}$$

is below a threshold, which we denote by ζ. The expression $|\cdot|$ indicates the sum of absolute values. The normalization with the division by $|\Omega|$ yields that the stopping criterion is independent of the size of the image domain Ω.

3.7.3 Error Measures

In order to be able to compare the performances of the different models in terms of numbers, we introduce two of the most common used error measures for optical flow, the absolute endpoint error (AEE) and the angular error (AE). These errors are amongst others the basis for the evaluation in the Middlebury database (Middlebury (2016)), which is one of the most popular databases for motion estimation models, cf. Baker et al. (2011). The AEE was first introduced in Otte and Nagel (1994). As the name implies, it measures the absolute difference between two flow fields, which are usually the calculated flow field (v^1, v^2) and the ground truth flow (v^1_{GT}, v^2_{GT}). For nPx

identifying the number of pixels in a single image frame, the AEE is defined as

$$AEE := \frac{1}{nPx} \sum_{i=1}^{nPx} \sqrt{(v^1(i) - v^1_{GT}(i))^2 + (v^2(i) - v^2_{GT}(i))^2}.$$

The AE was originally introduced in Fleet and Jepson (1990) and became popular through the survey in Barron et al. (1994). The idea is to measure the angles between two normalized velocity fields $(\hat{v}^1, \hat{v}^2)$ and $(\hat{v}^1_{GT}, \hat{v}^2_{GT})$. Since in case of zero motion the normalization would include a division by zero, the velocity fields are projected into the three-dimensional space, i.e.

$$\hat{\boldsymbol{v}} := \frac{(v^1, v^2, 1)}{\sqrt{\|\boldsymbol{v}\|^2 + 1}}, \qquad \hat{\boldsymbol{v}}_{GT} := \frac{(v^1_{GT}, v^2_{GT}, 1)}{\sqrt{\|\boldsymbol{v}_{GT}\|^2 + 1}}.$$

The angles between these normalized velocity fields are measured as

$$AE := \frac{1}{nPx} \sum_{i=1}^{nPx} \arccos(\hat{\boldsymbol{v}}(i) \cdot \hat{\boldsymbol{v}}_{GT}(i)).$$

3.8 Synthetic Results

The intention of the previously described optical flow models is to enable a real data application. Nevertheless, we need synthetic data sets to be able to test the models and to analyze their behavior based on error measures. We use the realistic data sets from the Middlebury database (Middlebury (2016)) as test data, since they combine a huge range of different characteristics. The names of the data sets we use are Rubber Whale, Hydrangea, and Dimetredon. All of them show different objects that are rotated and shifted between the frame. A rotation usually results in a smooth transition in the velocity field, whereas a shifting results in a piece wise constant movement. Thus, these data sets combine two of the most challenging characteristics. Noise is not included in any of the data sets.

3.8.1 Comparison of Data Fidelities

We start the evaluation of the different data fidelities by comparing only the linearized data fidelities $\mathcal{D}_{Lin1}$ and $\mathcal{D}_{Lin2}$, which we introduced in Section 3.2.1. We combine both of them with the standard regularizers $\mathcal{R}_{L^2}$ and $\mathcal{R}_{TV}$ (see Section 3.3.1). The original data sets from the Middlebury data base contain movement, which is larger than a single pixel. However, we already discussed that the linearized optical flow constraints are not reliable for large scale movement. This is why we downscaled the movement of

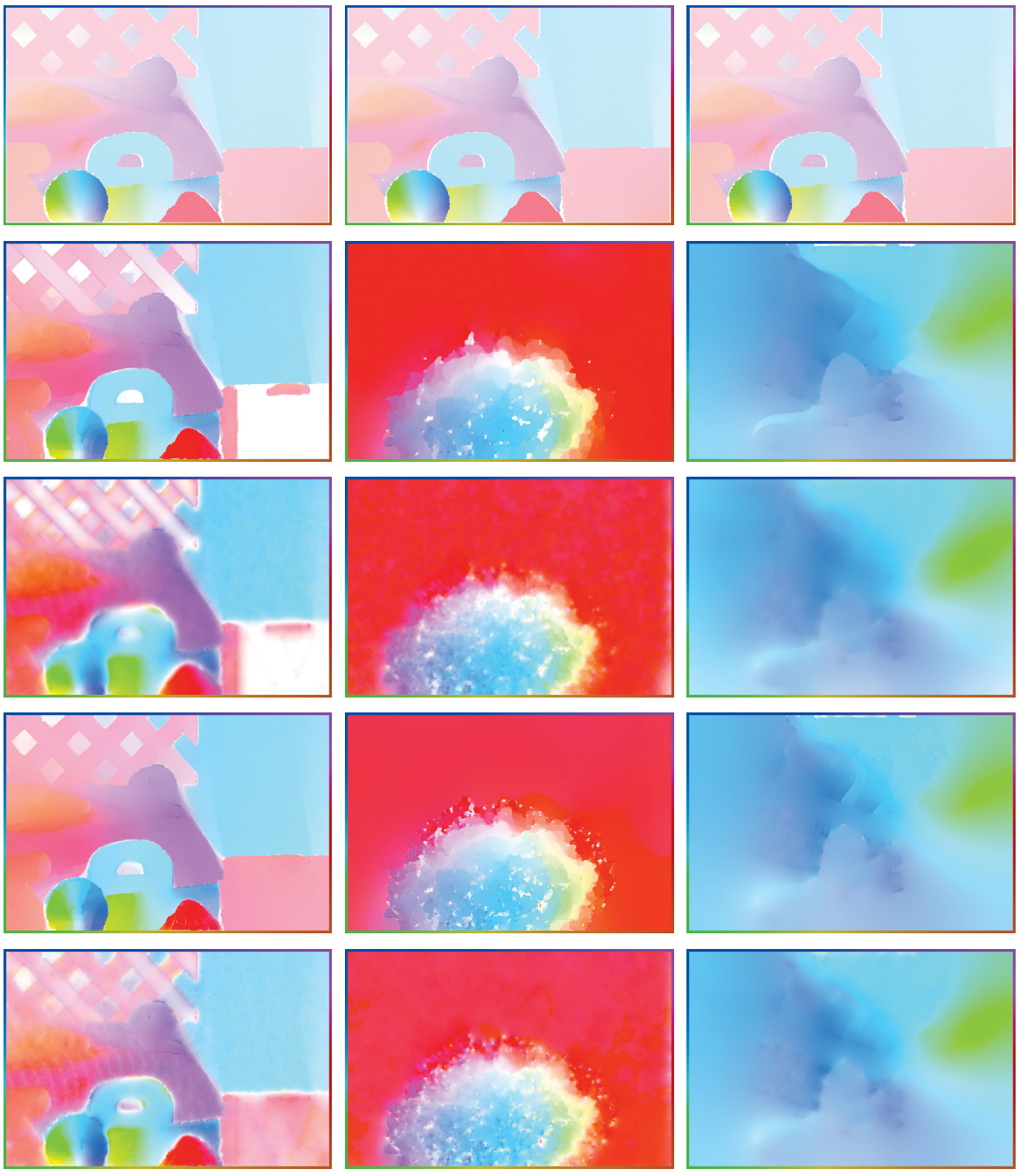

Figure 3.4: Comparison between the different data fidelities for the downscaled data sets Rubber Whale (left), Hydrangea (middle), and Dimetredon (right). Top to bottom: ground-truth, L^1-TV linear, L^1-L^2 linear, L^2-TV linear, L^2-L^2 linear.

	Rubber Whale		**Hydrangea**		**Dimetredon**	
	AEE	AE	AEE	AE	AEE	AE
L^1-TV	0.0295	0.0272	0.0342	0.0532	0.0392	0.0523
L^1-L^2	0.0368	0.0319	0.0345	0.0537	**0.0369**	**0.0504**
L^2-TV	**0.0188**	**0.0201**	**0.0299**	**0.0509**	0.0402	0.0530
L^2-L^2	0.0262	0.0248	0.0333	0.0530	0.0404	0.0528

Table 3.4: Absolute endpoint error (AEE) and angular error (AE) for results from Figure 3.4.

	linear		**nonlinear**	
	AEE	AE	AEE	AE
L^1-TV	**0.2219**	**0.0791**	0.0715	0.0346
L^1-L^2	0.2313	0.0857	0.0921	0.0430
L^2-TV	0.2263	0.0817	**0.0637**	**0.0319**
L^2-L^2	0.2303	0.0860	0.0836	0.0400

Table 3.5: Absolute endpoint error (AEE) and angular error (AE) for results from Figure 3.5.

the used data sets to a maximum of one single for this approach. Hence, the direction of movement is still the same as in the original data sets. Figure 3.4 shows the results of this comparison, as well as the ground truth velocity fields. Table 3.4 contains the particular error measures AEE and AE. The results show that the choice of the data fidelity is strongly dependent on the specific data set. Since the images do not contain any noise, the only criterion for the performance of the models is the movement itself. In the results it can be seen that the L^1 data fidelities have more problems with recognizing motion of a constant object, as it occurs with the red box in the lower right corner of the Rubber Whale data sets. Here, the L^1 data fidelity detects movement only at the outer edges of the object, whereas the L^2 data fidelity recognizes a movement of the whole object. Regarding the measured errors in Table 3.4 it becomes obvious that the performance of the models is also strongly dependent on the combination with the regularization. With some data sets a combination with a specific regularizer works better than with another one. We will further evaluate the influence of the regularizers in Section 3.8.2. Altogether, it is not possible to point out a single data fidelity that works best for all data sets.

We continue by comparing the linearized with the nonlinear optical flow constraints. For this purpose, we use the Rubber Whale data set, which we also used for the previous approach. However, in contrast to the downscaled movements in the previous setting,

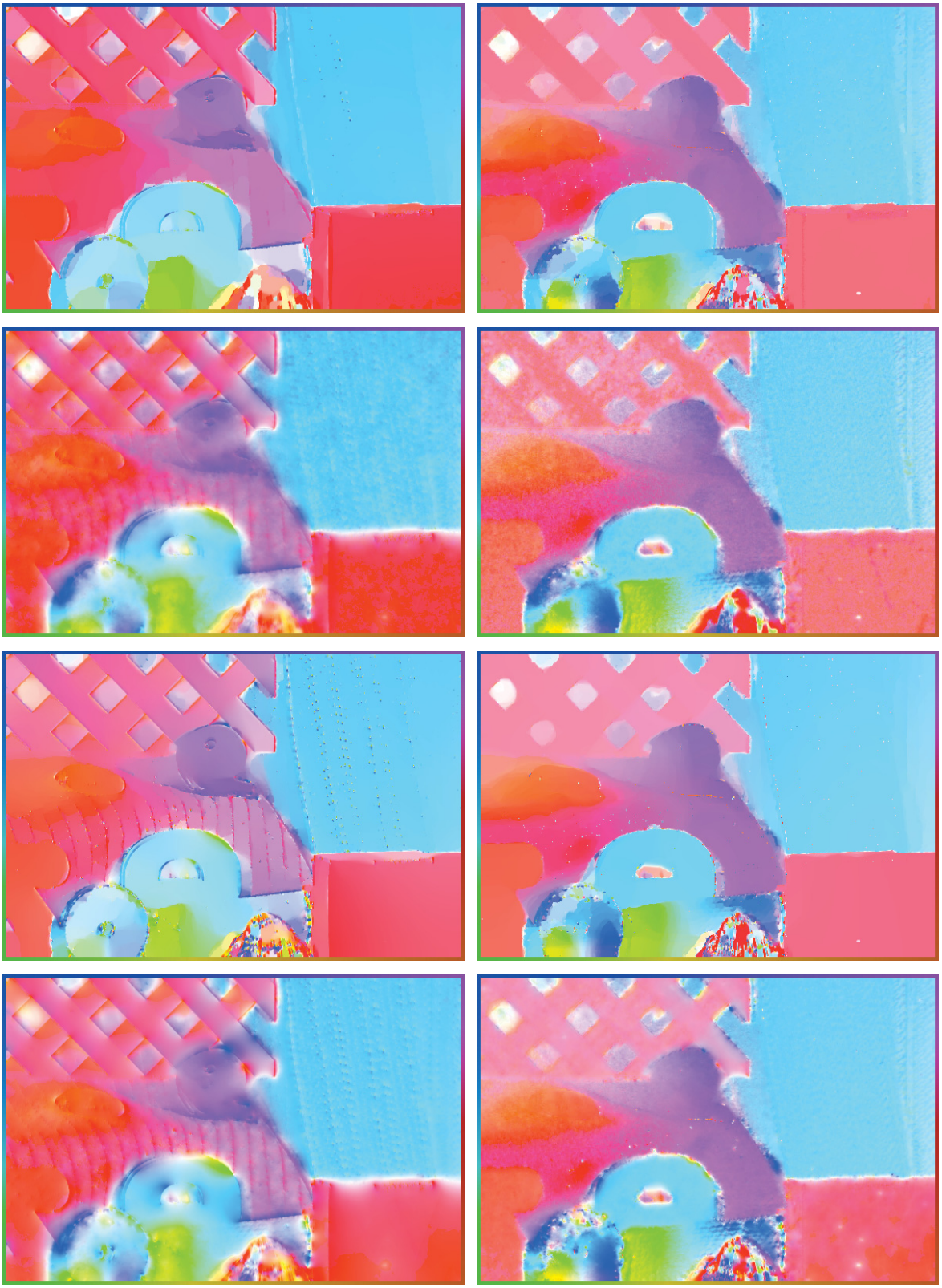

Figure 3.5: Comparison between the linearized and the nonlinear optical flow constraint for the Rubber Whale data set. Left: linearized data fidelities, right: nonlinear data fidelities. From top to bottom: L^1-TV, L^1-L^2, L^2-TV, L^2-L^2.

we use the original large scale movements for this comparison. For the evaluation we use the L^1-TV model, the L^1-L^2 model, the L^2-TV model and the L^2-L^2 model each in the linearized and in the nonlinear version. The resulting velocity fields can be found in Figure 3.5. A comparison of the visual results already reveals the main drawback of the linearized model. In areas with fast movement, like in the violet part in the center of the images, all of the linearized models show linear fragments, where there should actually be smooth or even constant movement. The nonlinear models solve this clearly better. None of them shows such characteristic fragments in the center area. However, smaller fragments are visible in the results of nonlinear models e.g. in the area of the rotating disk, especially for those models with an L^2 data fidelity. The particular error measures AEE and AE are summarized in Table 3.5. Here the difference between the linearized and the nonlinear models become even more obvious. Comparing both versions of each model shows that both measured errors are more than double the size for the linearized models than for the nonlinear models. Also the order in terms of best performance is different for both types of models, even though all approaches are based on the same data set. Regarding only the linearized models, the L^2-TV model works clearly better than the other ones. Concerning the nonlinear models, the L^2-TV model yields the best results. Altogether, the idea of not linearizing the optical flow constraint seems to work out well at least for the chosen data set.

3.8.2 Comparison of Regularizers

To test the performance of the different regularizers, we only use the nonlinear optical flow constraints $\mathcal{D}_{NL}$ that we derived in Section 3.2.2, to be able to produce reliable results for large scale motion. We use the original large scale variant of the Rubber Whale and the Hydrangea data set that we already used in the previous section. The results are shown in Figure 3.6 and the corresponding error measures are stated in Table 3.6. For these data sets both the visual appearance as well as the resulting errors are by far the worst for the squared L^2 norm used as regularization. Especially for the actually constant background of the Hydrangea data set, this regularizer is not able to produce suitable results. The TV and the TV/L^2 regularization for this data set are considerably superior to the squared L^2 norm, and nearly identical to each other. Regarding the errors, the TV model performs slightly better. The result for the TV/TV model for this data set appears similar to the other TV-based regularizers. Nevertheless, the errors are a little worse. The reason for this is that the Hydrangea data set does not really have smooth parts. It consists of lots of very fine structures, which can be recognized as smooth parts by the extended regularizers. The pure TV model also produces edges in regions where the structures are so fine that the other

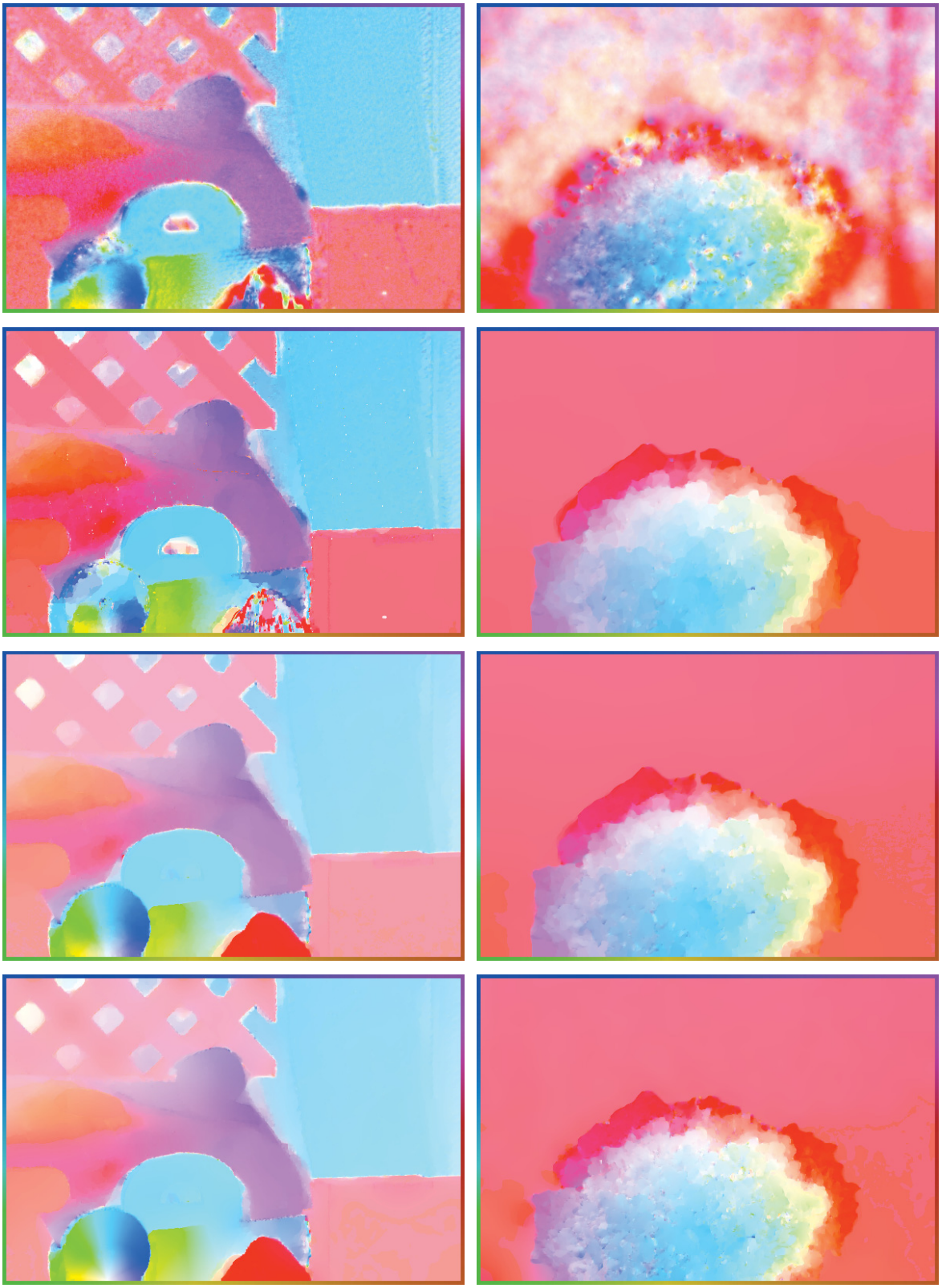

Figure 3.6: Comparison between the different regularizers for the Rubber Whale data set (left) and the Hydrangea data set (right). From top to bottom: L^1-L^2, L^1-TV, L^1-TV/L^2, L^1-TV/TV.

	Rubber Whale		Hydrangea	
	AEE	AE	AEE	AE
L^1-L^2	0.0921	0.0430	0.1228	0.0473
L^1-TV	0.0715	0.0346	**0.0878**	**0.0443**
L^1-TV/L^2	**0.0660**	**0.0322**	0.0881	0.0444
L^1-TV/TV	0.0664	0.0327	0.0955	0.0453

Table 3.6: Absolute endpoint error (AEE) and angular error (AE) for results from Figure 3.6.

models calculate them as smooth areas. This is a specific advantage in connection with the Hydrangea data set. For the Rubber Whale data set, the TV/L^2 and the TV/TV model yields almost identical results, which are clearly better than the L^2 and the TV results. Both nicely produce the smooth transitions without any staircasing effect. This becomes obvious especially in the area of the rotating disk, where the flow field is very smooth. Nevertheless, they also depict perfectly sharp edges. The only part, which both of them are not able to recover reasonably, is the inner part of the letter *e*. Analyzing the error measures it turns out that the result of the TV/L^2 regularizer is even slightly better than of the TV/TV regularizer. However, the differences are insignificant. In general, the results reveal that the extended regularizers outperform the standard regularizers if they are applied to data sets that actually feature both smooth parts and sharp edges.

3.8.3 Influence of Bregman Iterations

In this section we show the influence of the Bregman iterations for the L^2-TV optical flow models. We analyze the linear, as well as the nonlinear variant. One of the advantages of Bregman iterations is that the result is not as strongly dependent on the regularization parameter as for models without Bregman iterations. The parameter has more the effect of a step size. The larger the regularization parameter is, the more Bregman iterations are required to achieve the best result. Hence, it is adequate to simply choose a fixed parameter and calculate as many Bregman iterations as necessary. Since during the Bregman iterations the contrast and thus also the structures are enhanced, the regularization parameter has to be chosen considerably smaller than for calculations without Bregman iterations. This means that the result is calculated too smooth in the beginning, which is however balanced during the iterations.

Figure 3.7 shows the results for 1, ..., 10 Bregman iterations of the linearized L^2-TV optical flow model for the Rubber Whale data set. We previously verified that the

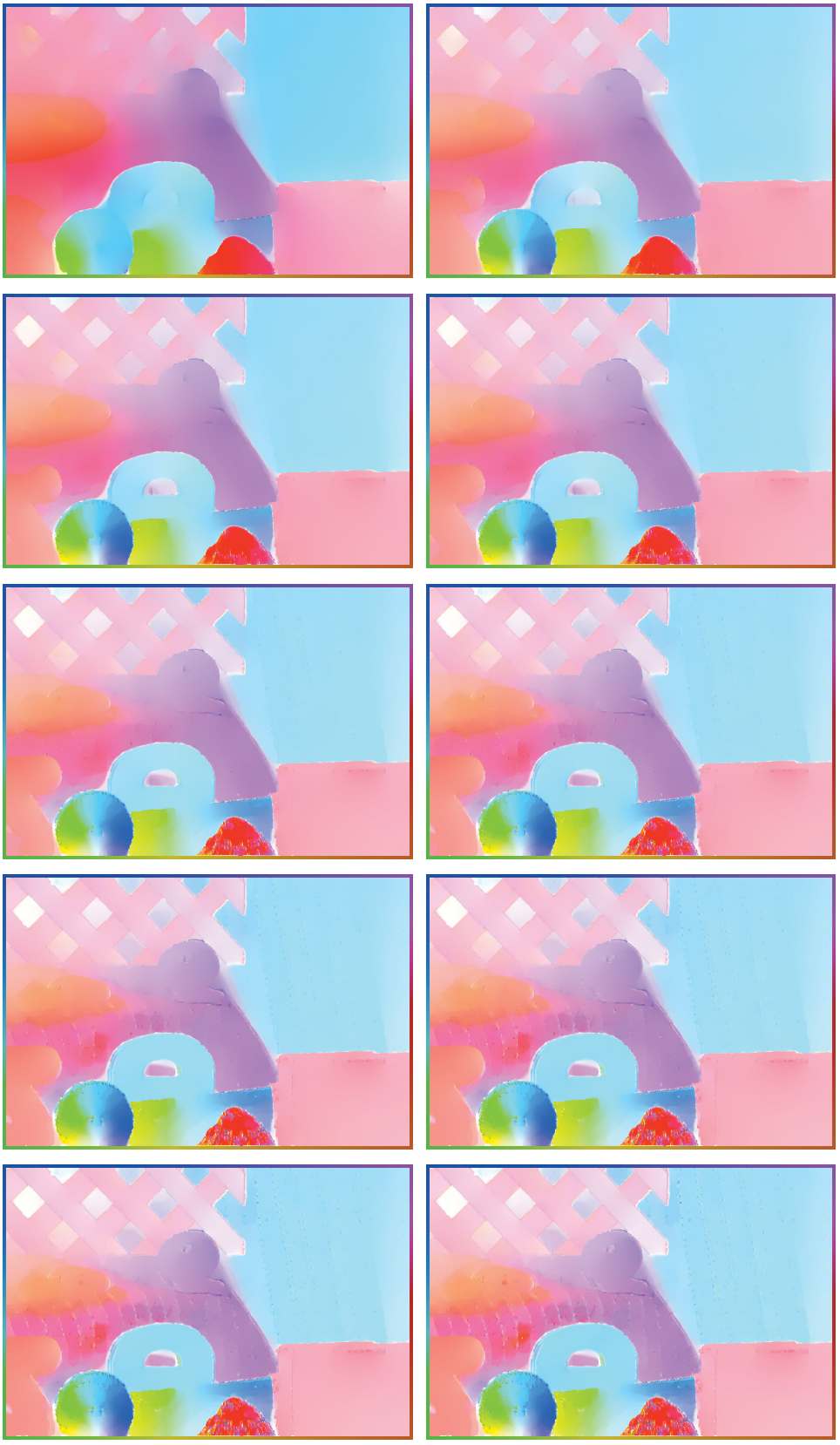

Figure 3.7: Results of the linearized L^2-TV model with Bregman iterations. From top left to bottom right: result after 1, ..., 10 Bregman iterations.

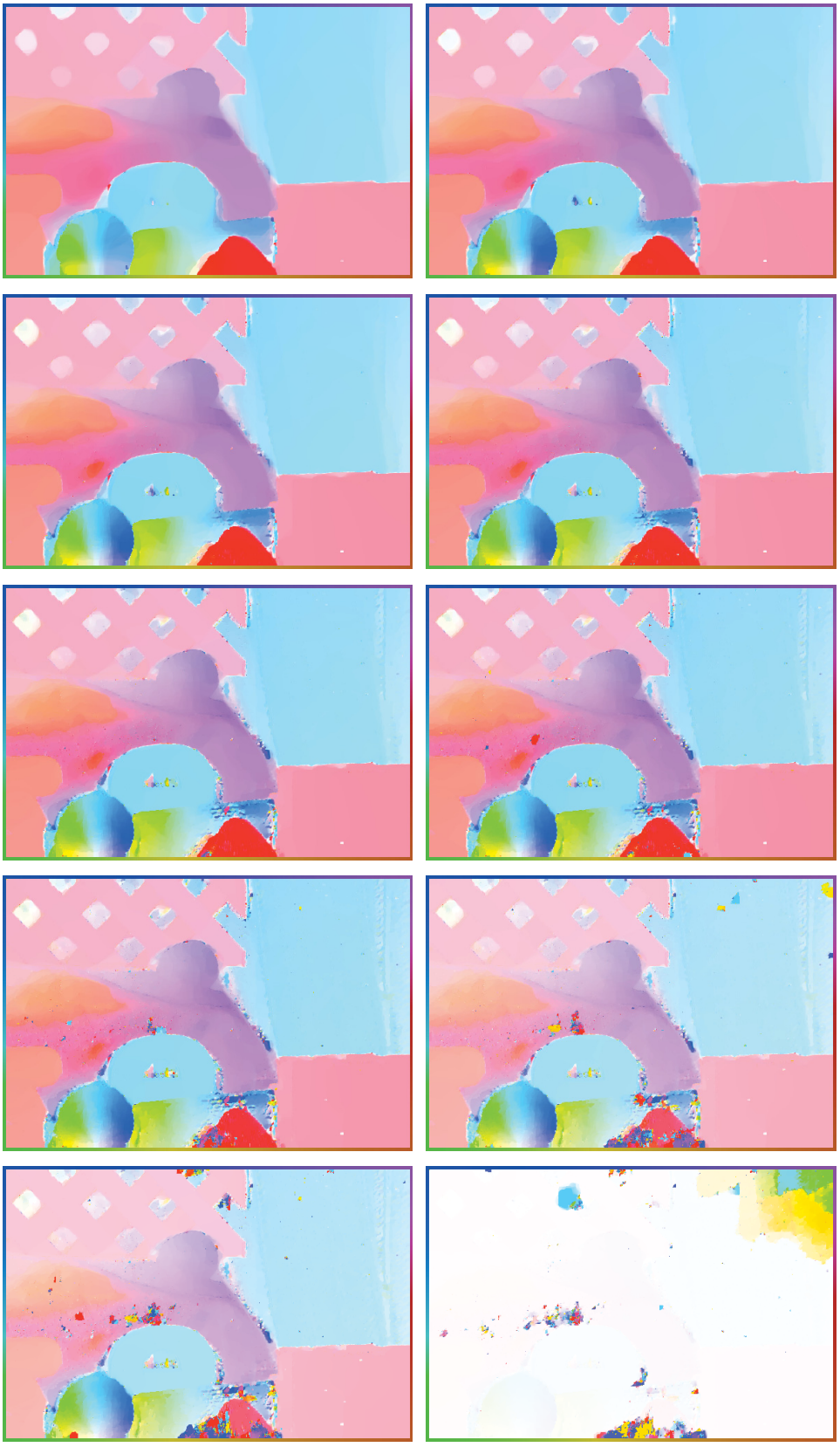

Figure 3.8: Results of the nonlinear L^2-TV model with Bregman iterations. From top left to bottom right: result after 1, ..., 10 Bregman iterations.

	linear		nonlinear	
#Breg.	AEE	AE	AEE	AE
1	0.0257	0.0246	0.0807	0.0367
2	0.0219	0.0220	0.0625	0.0313
3	**0.0215**	**0.0217**	**0.0617**	**0.0309**
4	**0.0215**	**0.0217**	0.0682	0.0313
5	0.0216	**0.0217**	9.1888	0.0326
6	0.0219	0.0219	20.9818	0.0342
7	0.0225	0.0222	171.2795	0.0366
8	0.0233	0.0227	442.8644	0.0427
9	0.0240	0.0232	506.1272	0.0452
10	0.0247	0.0236	512.7571	0.0904

Table 3.7: Absolute endpoint error (AEE) and angular error (AE) for results from Figure 3.7 and Figure 3.8.

linearized data fidelities are not able to produce accurate results for large scale motion. For this reason, we choose the downscaled data set with the small motion for this approach. The first iteration of the results is very smooth due to the choice of the regularization parameter. With every iteration step, the smaller structures become more and more visible. However, after about 6 iterations, there are also false structures visible, which are not existent in the ground truth flow field. These structures are even more enhanced with the following iterations. They are especially visible in the areas where there should actually be smooth transitions. The particular errors, which are stated in Table 3.7 document this effect. The best results are achieved for three and four Bregman iterations. With more iterations the errors become larger. This shows that it is necessary to stop the calculations after a distinct amount of iterations in order to achieve accurate results. For synthetic data sets with ground truth flow field, the most convenient way is to simply stop after the errors start to increase.

The results for the nonlinear L^2-TV model are presented in Figure 3.8. Again, we show the flow fields after 1, ..., 10 Bregman iterations for the Rubber Whale data set. Though, in this case we use the original large scale variant of the data set, in order to deal with a more realistic setting for the nonlinear model. Similar to the linearized variant, the result after the first Bregman iteration is very smooth as a consequence of the chosen regularization parameter. In the subsequent iterations the structures of the flow field become more obvious. After five or more Bregman iterations, the flow exhibits false artifacts. These occur as strong movements that point in completely false directions. The relation of these areas is slowly growing with the number of Bregman iterations. After nine iterations, the general scenery is still well recognizable. However, after ten

iterations the intensity of the false artifacts is extremely large, so that the scenery vanishes due to the scaling. Regarding the errors in Table 3.7, the best result is achieved after 3 Bregman iterations. After four iterations the errors are still relatively small. From the fifth iteration on, where also the first artifacts appear, the errors start to grow very fast. Consequently, for the nonlinear optical flow model with Bregman iterations it is even more important than for the linearized variant to stop the calculations at a certain point. Cf. Burger et al. (2007b) for a detailed analysis of the convergence of Bregman iterations.

4 Indirect Tracking

This chapter combines the previously introduced optical flow model with reconstruction tasks. We develop a model that reconstructs images at different time steps and simultaneously estimates the flow in between. An application for the derived model can be found in Chapter 6.

In the beginning of this chapter we give a short introduction to the general topic of indirect tracking and review possibilities for its numerical solution. Section 4.2 is devoted to reconstruction, which is necessary for the tracking. We especially focus on time-dependent reconstruction tasks. After introducing a time-dependent reconstruction operator, we list some examples for different types of time dependency. Subsequently, we present a very general model for indirect tracking, as well as four specific models. We also prove the existence of solutions for two of the specific models. In the final section we propose a computational approach for these models. We suggest an alternating minimization approach and explain the detailed implementation of the two subproblems, which handle the reconstruction and the motion estimation part. Since the model

4.1 Introduction

The motion estimation models we introduced in the previous chapter are all based on the assumption that the images at all time steps are directly available. The images and their corresponding intensities are the foundation for the derived optical flow constraints. However, in many image processing applications the available data are not yet perfect images. Hence, before being able to estimate motion, the images have to be reconstructed. In medical applications this the case for e.g. X-ray tomography, magnetic resonance imaging (MRI), positron emission tomography (PET) and many more. In

some cases the images are basically available, but they are either noisy, which is actually the case in most real data applications, or certain parts are missing completely. All those examples have in common that motion estimation based on the raw data does not lead to proper results, or is not possible at all.

Probably the most intuitive way to overcome this problem is to first reconstruct the images at every time step by applying e.g. Radon transform inversion, denoising, or inpainting algorithms, and afterwards estimating the motion based on the results. This way the reconstruction is used to obtain a proper velocity field. However, this procedure completely neglects the aspect that the velocity field might also help to improve the reconstruction, which is of great importance for the motion estimation. A more advanced method to obtain proper data for estimating motion is to use a single model for reconstruction and motion estimation. Recent results show that an indirect model can clearly enhance both the reconstruction and the estimated motion, see Dirks (2015) and Suhr (2015). The idea behind this kind of models is to alternately reconstruct the images and estimate the motion. In each of these steps, the information of the result of the previous step is included. This way the velocity field helps to improve the reconstruction, and the reconstruction helps to improve the velocity field. For such cases where the velocity field only serves as a supporting factor for the reconstruction, these models show great success (see Burger et al. (2016a)).

However, all of these models are based on the assumption that it is possible to reconstruct single images at every time step without relying on the data that belongs to the remaining time steps. This is e.g. the case for denoising of microscopy data. Here, every time step can be reconstructed separately. Even though the velocity field is considered as a priori information, the reconstruction itself remains independent of the time. Nevertheless, there are some reconstruction tasks where we have to deal with the problem that the temporal aspect is not only supporting, but absolutely mandatory. Considering e.g. continuous movement during a PET or single-photon emission computed tomography (SPECT) scan, a separate reconstruction will not result in adequate images for single time steps. Usually, the previously mentioned scanning methods rely on a certain amount of data, which is necessary for a proper reconstruction. This can only be collected over a distinct period of time. Regarding only the information that exists for a single time step is simply not sufficient. As soon as movement comes into play, the linkage between the information gained at different time steps becomes a challenging task.

The main goal of this chapter is to introduce and to analyze an indirect model for

motion estimation and time-dependent image reconstruction. For this purpose, we build on the models of Dirks (2015) and Suhr (2015) and adapt them to a time-dependent reconstruction setting. The combination with flow estimation is supposed to serve as the missing link between the different time steps and thus allows a reconstruction that would not have been possible on its own.

4.2 Time-Dependent Reconstruction

In this section we focus on the reconstruction, which serves as a basis for indirect tracking. We concentrate on time-dependent types of reconstructions, since in these cases the interaction between tracking and reconstruction is of particular importance.

4.2.1 Time-Dependent Reconstruction Operator

Concerning the mathematical model, the main difference between the cases where a reconstruction relies only on a single time step and those where it relies on multiple time steps is in the reconstruction operator, which we denote by H. The exact form of H depends on the type of reconstruction, i.e. whether it is supposed to denoise an image, to inpaint it, or to reconstruct it from e.g. tomography data. In order to reconstruct the unknown image u, with $u : \Omega \times [0, T] \to \mathbb{R}^+$, it is necessary to solve the inverse problem

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

where f is the given measured data. If the reconstruction is only dependent on a single time step, the operator H is time-independent. If on the other hand information of several time steps is required, H is dependent on time, i.e. we have to deal with a nonlinear operator of the type $H(t)$ with $t \in [0, T]$. Consequently, instead of (4.1) the inverse problem we have to solve is

$$H(t)u(\cdot, t) = f(\cdot, t). \tag{4.2}$$

Only as long as there is no motion in the measured data, the time-dependency can obviously be neglected. This is the reason why reconstruction based on static measurements is considerably less challenging compared to reconstruction based on dynamic measurements. Without further a priori knowledge, it is typically not possible to connect the information measured at different time steps in a useful way. It is however necessary if the reconstruction of an image at a certain time step requires the information of the remaining measurements.

In general, the reconstruction operator can be modeled to map onto arbitrary Hilbert spaces $Y(t)$ for each time step, i.e.

$$H(t) : L^p(\Omega) \to Y(t),$$

with $p \in \{1, 2\}$. Only in case that all measurements are available at each time step, this mapping can be linear. The general data fidelity, which approximates the solution of the inverse problem (4.2), is given by

$$\mathcal{D}(u, f) = \int_0^T \|H(t)u(\cdot, t) - f(t)\|^p_{Y(t)}\, dt,$$

with $u \in L^p([0, T]; L^2)$.

4.2.2 Examples for Time-Dependency

This section gives different examples for applications with required time-dependent reconstruction.

- A first example for time-dependency is a given measurement, where some of the time steps are missing. In this case the reconstruction operator is of the form

$$H(t) = \begin{cases} I, & \text{for measured time steps} \\ 0, & \text{for missing time steps} \end{cases},$$

  where I is an operator, which is only applied to the data of the actually measured time steps, e.g. the identity. Thus, H defines an operator on the complete set of time-steps even though some of them are not available.

- Similar to the missing time steps in the first example, time-dependent reconstruction can also be used for super-resolution in time. In the same way as above, several time steps can be added to the ones that are already given. For the reconstruction operator, this means

$$H(t) = \begin{cases} S, & \text{for given time steps} \\ 0, & \text{for additional time steps} \end{cases},$$

  where S denotes a reconstruction operator, which is applied to the given measurements. This proceeding acts like a reduction of the intervals between the time

steps. Usually, this goes along with a reduced velocity of the measured object and can therefore improve the motion estimation result.

- A third example is given by measurements, where the values of some pixels are false or missing in certain time steps. Then for a single time step the image domain is only a subset of the complete domain Ω, i.e. $\Omega(t) \subset \Omega$. Consequently, the Hilbert space is given by

$$Y(t) = L^2(\Omega(t)).$$

  The reconstruction can therefore be identified with an inpainting operator that fills in values at the missing pixels.

- A stronger variant of the idea of the previous example is given in compressed sensing. In this application the goal is to reconstruct data based on only a sparse series of sample measurements. During this process the amount of pixels is being drastically reduced in a random manner in every step. Again, for the given Hilbert space it holds

$$Y(t) = L^2(\Omega(t)),$$

  where $\Omega(t) \subset\subset \Omega$ is a random subset.

4.3 Models for Indirect Tracking

Since we intend to combine two different imaging tasks, the indirect model contains two data fidelities $\mathcal{D}_{rec}(u, f)$ and $\mathcal{D}_{mot}(u, \boldsymbol{v})$. The first one handles the reconstruction, and the second one handles the motion estimation. We already discussed in Chapter 3 that motion estimation is an ill-posed problem, which is why we add a regularizer $\mathcal{R}_{\boldsymbol{v}}(\boldsymbol{v})$ for the velocity field. The same is true for a wide range of reconstruction tasks. Consequently, we complement the model with a regularizer $\mathcal{R}_u(u)$ for the image intensity. Altogether, we want to solve the following variational model:

Model 4.1. (General Model for Indirect Tracking)

$$\min_{u,\boldsymbol{v}} \mathcal{D}_{rec}(u, f) + \alpha\mathcal{R}_u(u) + \beta\mathcal{R}_{\boldsymbol{v}}(\boldsymbol{v}) + \gamma\mathcal{D}_{mot}(u, \boldsymbol{v}).$$

The parameters α, β and γ serve as weighting parameters between the different terms. Since there are different options for every single term of this model, there is consequently

a wide range of possible combinations. For the sake of simplicity we restrict ourselves to a more specific model for the following discussion.

4.3.1 Specific Models

The first data fidelity $\mathcal{D}_{rec}(u, f)$ of Model 4.1 is supposed to solve the reconstruction problem (4.2), which is why we want to minimize $Hu - f$, equipped with an appropriate norm. There are various possibilities for the choice of the norm. The suitability strongly depends on the type of noise in the data (see Section 3.2.1). For the sake of simplicity we restrict ourselves to two different cases, which are the L^1 norm and the squared L^2 norm. For the first case we set

$$\mathcal{D}_{rec}^{L^1}(u, f) = \int_0^T \|H(t)u(\cdot, t) - f(\cdot, t)\|_1 \, dt,$$

and for the second case

$$\mathcal{D}_{rec}^{L^2}(u, f) = \int_0^T \frac{1}{2}\|H(t)u(\cdot, t) - f(\cdot, t)\|_2^2 \, dt.$$

The additional data fidelity $\mathcal{D}_{mot}(u, \boldsymbol{v})$ calculates the velocity field. Therefore, we use an optical flow constraint for this term. In principle, each of the data fidelities that we derived in Section 3.2 would be a suitable choice. However, for the norm of this data fidelity we restrict ourselves to L^1, since in many cases it produces better results than the squared L^2 norm. Concerning the linearity, we consider both the linearized and the nonlinear variants. In the notation of Section 3.2.2 we are dealing with two consecutive image frames u_0 and u_1. In this case the frame u_0 is equivalent to the frame u in this section. For this reason, we proceed writing u instead of u_0. Thus, our suggestions for $\mathcal{D}_{mot}(u, \boldsymbol{v})$ are

$$\mathcal{D}_{mot}^{Lin}(u, \boldsymbol{v}) = \|\boldsymbol{v} \cdot \nabla u + \partial_t u\|_1,$$

and

$$\mathcal{D}_{mot}^{NL}(u, \boldsymbol{v}) = \|u_1 - u(\boldsymbol{v})\|_1.$$

Concerning the regularizer $\mathcal{R}_{\boldsymbol{v}}(\boldsymbol{v})$ for the velocity field, again all of the alternatives that

we introduced in Section 3.3 can be used for the indirect model. For the specific models that we want to analyze in this chapter, we use the TV norm, since it is appropriate for many different applications:

$$\mathcal{R}_{\boldsymbol{v}}(\boldsymbol{v}) = \|\nabla \mathbf{v}\|_1.$$

Concerning the choice of the regularizer $\mathcal{R}_u(u)$, there is also a wide range of possibilities. Most issues that we discussed for the regularizer for a velocity field $\boldsymbol{v}$ are easily transferable to the regularization of the image intensity u. For the same reason as above, we focus on the TV norm and set

$$\mathcal{R}_u(u) = \|\nabla u\|_1.$$

Inserting $\mathcal{D}_{rec}^{L^1}$ for the reconstruction as well as $\mathcal{D}_{mot}^{Lin}$ for the motion estimation into (3.3), we end up with the first specified model:

Model 4.2. (Linear L^1-TV-TV-L^1 Model)

$$\begin{aligned}\arg\min_{u,\mathbf{v}} \int_0^T \|H(t)u(\cdot,t) - f(\cdot,t)\|_1 + \alpha\|\nabla u(\cdot,t)\|_1 + \beta\|\nabla \mathbf{v}(\cdot,t)\|_1 \\ + \gamma\|\boldsymbol{v}(\cdot,t)\cdot\nabla u(\cdot,t) + \partial_t u(\cdot,t)\|_1\, dt\end{aligned}$$

Using $\mathcal{D}_{rec}^{L^2}$ instead of $\mathcal{D}_{rec}^{L^1}$, we obtain the second specified model:

Model 4.3. (Linear L^2-TV-TV-L^1 Model)

$$\begin{aligned}\arg\min_{u,\mathbf{v}} \int_0^T \frac{1}{2}\|H(t)u(\cdot,t) - f(\cdot,t)\|_2^2 + \alpha\|\nabla u(\cdot,t)\|_1 + \beta\|\nabla \mathbf{v}(\cdot,t)\|_1 \\ + \gamma\|\boldsymbol{v}(\cdot,t)\cdot\nabla u(\cdot,t) + \partial_t u(\cdot,t)\|_1\, dt\end{aligned}$$

For $\mathcal{D}_{rec}^{L^1}$ in combination with with the nonlinear optical flow constraint $\mathcal{D}_{mot}^{NL}$ we achieve

Model 4.4. (Nonlinear L^1-TV-TV-L^1 Model)

$$\begin{aligned}\arg\min_{u,u_1,\mathbf{v}} \int_0^T \|H(t)u(\cdot,t) - f(\cdot,t)\|_1 + \alpha\|\nabla u(\cdot,t)\|_1 + \beta\|\nabla \mathbf{v}(\cdot,t)\|_1 \\ + \gamma\|u_1 - u\left(\boldsymbol{v}(\cdot,t),t\right)\|_1\, dt\end{aligned}$$

For the last specified model we combine $\mathcal{D}_{rec}^{L^2}$ with $\mathcal{D}_{mot}^{NL}$:

Model 4.5. (Nonlinear L^2-TV-TV-L^1 Model)

$$\begin{aligned}\arg\min_{u,u_1,\mathbf{v}} \int_0^T \frac{1}{2}\|H(t)u(\cdot,t) - f(\cdot,t)\|_2^2 + \alpha\|\nabla u(\cdot,t)\|_1 + \beta\|\nabla \mathbf{v}(\cdot,t)\|_1 \\ + \gamma\|u_1 - u\left(\boldsymbol{v}(\cdot,t),t\right)\|_1\, dt\end{aligned}$$

In both of the nonlinear models the variable u_1 is independent of t.

4.3.2 Existence of Minimizers

For the existence analysis we reformulate the previously defined models. We extract the optical flow constraint and use it as a hard constraint for the remaining functional. Unfortunately, we are not able to prove the existence of minimizers for the variants with nonlinear optical flow constraint, at least for the L^1 case. Therefore, we only regard the linearized versions in this section. Hence, for Model 4.2 we obtain the main functional

$$J_1(u,\boldsymbol{v}) = \int_0^T \|H(t)u(\cdot,t) - f(\cdot,t)\|_1 + \alpha|u(\cdot,t)|_{BV}^p + \beta|\boldsymbol{v}(\cdot,t)|_{BV}^q\, dt, \tag{4.3}$$

with $1 < p, q \in \mathbb{R}$. Analogously, for Model 4.3 we achieve

$$J_2(u,\boldsymbol{v}) = \int_0^T \frac{1}{2}\|H(t)u(\cdot,t) - f(\cdot,t)\|_2^2 + \alpha|u(\cdot,t)|_{BV}^p + \beta|\boldsymbol{v}(\cdot,t)|_{BV}^q\, dt. \tag{4.4}$$

We assume the functionals to be smaller or equal to a constant c. In both cases the constraint is given by

$$\|r\|_{\mathcal{M}(\Omega\times[0,T])} \leq R(c), \qquad \text{with } r(\cdot,t) \coloneqq \boldsymbol{v}(\cdot,t)\cdot\nabla u(\cdot,t) + \partial_t u(\cdot,t)$$

and $\mathcal{M}$ being a Radon measure. For $R(c) = 0$, the constraint indicates $\boldsymbol{v}\cdot\nabla u + \partial_t u = 0$.

To show the existence of solutions we use the fundamental theorem of optimization (Theorem 2.1), i.e. we need to show

- coercivity of the functionals $J_1(u,\boldsymbol{v})$ and $J_2(u,\boldsymbol{v})$,
- lower semi-continuity of the functionals $J_1(u,\boldsymbol{v})$ and $J_2(u,\boldsymbol{v})$,
- closedness of the constraint $\boldsymbol{v}_n\cdot\nabla u_n + \partial_t u_n \to \boldsymbol{v}\cdot\nabla u + \partial_t u$ for $u_n \to u$ and $\boldsymbol{v}_n \to \boldsymbol{v}$ in suitable spaces.

The constraint $\mathcal{D}_{mot}(u_n,\boldsymbol{v}_n)$ is independent of the main functionals $J_1(u,\boldsymbol{v})$ and $J_2(u,\boldsymbol{v})$, and can therefore be treated for both functionals simultaneously. For the remainder

of this section we assume that $H : L^1(\Omega \times [0,T]) \to L^1(\Omega \times [0,T])$ is a continuous operator.

In order to prove coercivity for the functional $J_1(u, \boldsymbol{v})$, we define the subspace BV_0 as follows:

$$BV_0 = \left\{ u \in BV(\Omega) \middle| \int_\Omega u \, dx = 0 \right\}.$$

With this definition we are able to split every $u \in BV(\Omega)$ into its mean value and a part in $BV_0(\Omega)$, i.e.

$$u = \bar{u} + u_0, \quad \forall u \in BV(\Omega),$$

with $\bar{u}$ being the mean value of u, and $u_0 \in BV_0(\Omega)$. Since we want to use the theorem of Banach-Alaoglu (Theorem 2.2) to prove the coercivity, we need to define a space whose dual is given by $BV_0(\Omega)$. We stick to the proceeding in Burger and Osher (2013) and define the normed space

$$Z_0 = \left\{ \nabla \cdot g \middle| g \in C_0^\infty(\Omega; \mathbb{R}^d) \right\},$$

with the norm

$$\|p\|_Z = \inf_{g \in C_0^\infty, \nabla \cdot g = p} \|g\|_{L^\infty}.$$

With this definition $BV_0(\Omega)$ is the dual of

$$Z := \bar{Z}_0.$$

Furthermore, TV is an equivalent norm on $BV_0(\Omega)$ (see Burger and Osher (2013), Lemma 3.5).

We additionally define a space $\mathcal{X}$ that combines the required spaces for u and $\boldsymbol{v}$:

$$\mathcal{X} := \{(u, \boldsymbol{v}) \in L^p([0,T], BV(\Omega)) \times L^q([0,T], BV(\Omega))\}$$

with $1 < p, q \in \mathbb{R}$. To state necessary conditions for u and $\boldsymbol{v}$, we define a set on $\mathcal{X}$ as follows:

$$\mathcal{D} := \left\{ (u, \boldsymbol{v}) \in \mathcal{X} \middle| \|\boldsymbol{v}\|_\infty \leq c_v, \, \|\nabla \cdot \boldsymbol{v}\|_r \leq c_g \right\},$$

with $r \leq q$. Finally, we are able to show the coervivity of $J_1(u, \boldsymbol{v})$:

Lemma 4.1. (Coercivity of $J_1(u, \boldsymbol{v})$)
Let $1 < p, q \in \mathbb{R}$, $u \in L^p([0,T]; BV(\Omega))$ and $\boldsymbol{v} \in L^q([0,T]; BV(\Omega))$. For $(u, \boldsymbol{v})$ in the set

$$\mathcal{S}(c) = \left\{(u, \boldsymbol{v}) \in \mathcal{D} \middle| J_1(u, \boldsymbol{v}) \leq c,\, \boldsymbol{v} \cdot \nabla u + \partial_t u = 0\right\},$$

it holds

$$\|u\|_{L^p([0,T];BV(\Omega))} \leq c, \qquad \text{and} \qquad \|\boldsymbol{v}\|_{L^q([0,T];BV(\Omega))} \leq c.$$

Then, $\mathcal{S}(c)$ is non-empty and compact in the weak* topology.

Proof. With the definition of the subspace BV_0, we split the functional $J_1(u, \boldsymbol{v})$ into

$$\begin{aligned} J_1(u, \boldsymbol{v}) =& J_1(\bar{u} + u_0, \boldsymbol{v}) \\ \leq& \int_0^T \|H(\bar{u} + u_0) - f\|_1 + \alpha \left(|\bar{u}(\cdot, t)|_{BV}^p + |u_0(\cdot, t)|_{BV_0}^p\right) \\ &+ \beta \left(|\bar{\boldsymbol{v}}(\cdot, t)|_{BV}^q + |\boldsymbol{v}_0(\cdot, t)|_{BV_0}^q\right) dt. \end{aligned}$$

With the previous definition of the Banach space Z, we can directly achieve bounds $c_u \geq |u_0(\cdot, t)|_{BV_0}^p$ and $c_{\boldsymbol{v}} \geq |\boldsymbol{v}_0(\cdot, t)|_{BV_0}^p$. Additionally, it is clear that

$$|\bar{u}(\cdot, t)|_{BV}^p = |\bar{\boldsymbol{v}}(\cdot, t)|_{BV}^q = 0$$

and thus these terms are also bounded. We furthermore split the operator H into a part $\bar{H} : L^p([0,T]) \to \Omega$, which is applied to $\bar{u}$, and a part $H_0 : L^p([0,T]; BV_0) \to \Omega$, which is applied to u_0. With this formulation we obtain

$$\|H(\bar{u} + u_0) - f\|_1 = \|\bar{H}\bar{u} + H_0 u_0 - f\|_1.$$

Due to the positivity of the single terms and the bound c on the functional $J_1(u, \boldsymbol{v})$, we know that $\|\bar{H}\bar{u} + H_0 u_0 - f\|_1$ is also smaller or equal to c. With a triangle inequality we achieve

$$\begin{aligned} c &\geq \|\bar{H}\bar{u} + H_0 u_0 - f\|_1 \\ &\geq \|\bar{H}\bar{u}\|_1 - \|H_0 u_0\|_1 - \|f\|_1. \end{aligned}$$

Solving the inequality with respect to $\|\bar{H}\bar{u}\|_1$ yields

$$\|\bar{H}\bar{u}\|_1 \leq c + \|H_0 u_0\|_1 + \|f\|_1.$$

This gives us a bound $\bar{c} \geq \|\bar{u}\|_1$. For $p \geq 2$ we would also achieve a bound for $\|\bar{H}\bar{u}\|_p^p$. However, since $\bar{H} : \mathbb{R} \to \mathbb{R}$, we do not obtain boundedness in $L^1(\Omega)$. Hence, we have to access characteristics of the constraint

$$\|r\|_{\mathcal{M}(\Omega\times[0,T])} \leq R(c),$$

to obtain boundedness for $\|\bar{H}\bar{u}\|_1$. With the triangle inequality it holds

$$\|\partial_t u\|_{\mathcal{M}(\Omega\times[0,T])} \leq \int_0^T \int_\Omega |\boldsymbol{v} \cdot \nabla u|_1 \, dx \, dt + \int_0^T \int_\Omega |r|_1 \, dx \, dt.$$

Application of the splitting of u into u_0 and $\bar{u}$ yields

$$\begin{aligned}\int_0^T \int_\Omega |\partial_t u_0| \, dx \, dt + \int_0^T \int_\Omega |\partial_t \bar{u}| \, dx \, dt \leq \int_0^T \int_\Omega |\boldsymbol{v} \cdot \nabla u_0| \, dx \, dt \\ - \int_0^T \int_\Omega |\boldsymbol{v} \cdot \nabla \bar{u}| \, dx \, dt + \int_0^T \|r\|_1 \, dt.\end{aligned}$$

With the definition of the space BV_0, the first term on the left-hand side is equal to zero. The same applies for the second term on the right-hand side due to the gradient applied to a constant. Since $|\partial_t \bar{u}|$ is also constant in space, the integration for this term can be neglected. Hence, it remains

$$\int_0^T |\partial_t \bar{u}| \, dt \leq \int_0^T \int_\Omega |\boldsymbol{v} \cdot \nabla u_0| \, dx \, dt + \int_0^T \|r\|_1 \, dt.$$

We use the Hölder inequality to obtain

$$\int_0^T |\partial_t \bar{u}| \, dt \leq \int_0^T \|\boldsymbol{v}\|_\infty \|\nabla u_0\|_1 \, dt + \int_0^T \|r\|_1 \, dt.$$

Exploitation of the bound for $\|\boldsymbol{v}\|_\infty$ achieves

$$\int_0^T |\partial_t \bar{u}| \, dt \leq c_g^T \|u_0\|_{BV} + \int_0^T \|r\|_1 \, dt.$$

Consequently, $|\partial_t \bar{u}|$ is bounded due to the already investigated bound of $\|u_0\|_{BV}$ and the constraint $\int_0^T \|r\|_1 \, dt \leq R(c)$. We proceed by showing that this bound also induces

a bound for $\bar{u}$. For every $s \in \Omega$ we write

$$\begin{aligned}\bar{u}(t) &= \bar{u}(s) + \int_0^t \partial_t \bar{u}(\tau)\, d\tau \\ &= \frac{1}{T}\int_0^T \bar{u}(s)\, ds + \int_0^T \int_0^t \partial_t \bar{u}(\tau)\, d\tau\, ds.\end{aligned}$$

Using the absolute value we end up with the L^1 norm on the right-hand side:

$$\|\bar{u}\|_\infty \leq \frac{1}{T}\|\bar{u}\|_1 + \hat{c}\|\partial_t \bar{u}\|_{\mathcal{M}(\Omega\times[0,T])},$$

where $\hat{c}$ is a positive constant. With the bounds $\bar{c} \geq \|\bar{u}\|_1$ and $\tilde{c} \geq \|\partial_t \bar{u}\|_1$ we also obtain a bound

$$C \geq \|\bar{u}\|_\infty, \qquad \text{with } C \geq \frac{\bar{c}}{T} + \hat{c} + \tilde{c}.$$

Due to the continuity of the operator $\bar{H}$, this gives us boundedness of $\|\bar{H}\bar{u}\|_1$. Finally, every single term is bounded in $L^1(\Omega)$ and hence the set $\mathcal{S}(c)$ is compact. □

Lemma 4.2. (Lower Semi-Continuity of $J_1(u, \boldsymbol{v})$)
The functional $J_1(u, \boldsymbol{v})$ as defined in (4.3) is weak* lower semi-continuous in $\mathcal{D} \subset \mathcal{X}$.

Proof. We already showed that norms on Banach spaces are weak* lower semi-continuous (see Lemma 2.1). This statement also holds for affine norms as well as exponentials $p > 1$. Furthermore, sums of lower semi-continuous functionals are again lower semi-continuous. Consequently, since all summands of $J_1(u, \boldsymbol{v})$ are functionals on Banach spaces, we obtain weak* lower semi-continuity for the complete functional. □

It remains to show closedness of the constraint:

Lemma 4.3. (Closedness of the Constraint)
Let $(u, \boldsymbol{v} \in \mathcal{X})$. Furthermore, let $\|\boldsymbol{v}\|_\infty \leq c_\infty$ and $\nabla \cdot \boldsymbol{v} \in L^{p^* s}([0,T]; L^{2k}(\Omega))$ for $k, s > 1$ and p^* denoting the Hölder conjugate of p. Then the constraint

$$\int_0^T \|r(\cdot, t)\|_{\mathcal{M}(\Omega\times[0,T])}\, dt \leq R(c),$$

with $r(\cdot, t) = \boldsymbol{v}(\cdot, t) \cdot \nabla u(\cdot, t) + \partial_t u(\cdot, t)$, is closed.

Proof. For the case $R(c) = 0$ the convergence of the constraint in a distributional sense is shown in Burger et al. (2016a), Lemma 3.4. Hence, for $R(c) \neq 0$ we are able to find weakly convergent subsequences $u^n \in L^p([0,T]; \Omega)$ and $\boldsymbol{v}^n \in L^q([0,T]; \Omega)$ with

$$\boldsymbol{v}^n \cdot \nabla u^n + \partial_t u^n \rightharpoonup \boldsymbol{v} \cdot \nabla u + \partial_t u \qquad \text{in } \mathcal{D}(\Omega \times [0,T]).$$

Furthermore, there exist subsequences $u^{n_k} \in L^p([0,T];\Omega)$ and $\boldsymbol{v}^{n_k} \in L^q([0,T];\Omega)$ with

$$\boldsymbol{v}^{n_k} \cdot \nabla u^{n_k} + \partial_t u^{n_k} \rightharpoonup^* w \qquad \text{in } \mathcal{M}(\Omega \times [0,T])$$

for $w \in \mathcal{M}(\Omega \times [0,T])$. Due to the uniqueness of the limit, it holds $w = \boldsymbol{v} \cdot \nabla u + \partial_t u$. The same argument applies for every convergent subsequence in $\mathcal{M}(\Omega \times [0,T])$. Consequently, we achieve the following weak* convergence in $\mathcal{M}(\Omega \times [0,T])$:

$$\boldsymbol{v}^n \cdot \nabla u^n + \partial_t u^n \rightharpoonup^* \boldsymbol{v} \cdot \nabla u + \partial_t u$$

Since the norm in $\mathcal{M}(\Omega \times [0,T])$ is weak* lower semi-continuous, we end up with the boundedness of the constraint, i.e. $\|r(\cdot,t)\|_{\mathcal{M}(\Omega \times [0,T])} \leq R(c)$. □

Hence, we are able to show the existence of a minimal solution for the reformulated variant of Model 4.2:

Theorem 4.1. (Existence of Minimizers for the L^1-TV-TV-L^1 Model)
Let $1 < p, q$ and

$$J_1(u, \boldsymbol{v}) = \int_0^T \|Hu - f\|_1 + \alpha |u(\cdot,t)|_{BV}^p + \beta |\boldsymbol{v}(\cdot,t)|_{BV}^q \, dt.$$

Then, there exists a minimizer in the set

$$\mathcal{S}(c) = \left\{ (u, \boldsymbol{v}) \in \mathcal{D} \middle| J_1(u, \boldsymbol{v}) \leq c,\ \boldsymbol{v} \cdot \nabla u + \partial_t u = 0 \right\}.$$

Proof. The functional $J_1(u, \boldsymbol{v})$ fulfills the requirements for the fundamental theorem of optimization. We obtain coercivity for $J_1(u, \boldsymbol{v})$ from Lemma 4.1. The lower semi-continuity results from Lemma 4.2. The closedness of the constraints $\mathcal{D}_{mot}(u_n, \boldsymbol{v}_n)$ follows from Lemma 4.3. Consequently, there exists a minimizer for the reformulation of Model 4.2.

□

For the coercivity of $J_2(u, \boldsymbol{v})$, we use Dirks (2015), Lemma 5.3.3.:

Lemma 4.4. (Coercivity of $J_2(u, \boldsymbol{v})$)
For $1 < p, q \in \mathbb{R}$, let $u \in L^p([0,T]; BV(\Omega))$, $\boldsymbol{v} \in L^q([0,T]; BV(\Omega))$. For $(u, \boldsymbol{v})$ in the set

$$\mathcal{S}(c) = \left\{ (u, \boldsymbol{v}) \in \mathcal{D} \middle| J_2(u, \boldsymbol{v}) \leq c,\ \boldsymbol{v} \cdot \nabla u + \partial_t u = 0 \right\},$$

it holds

$$\|u\|_{L^p([0,T];BV(\Omega))} \leq c, \qquad \text{and} \qquad \|\boldsymbol{v}\|_{L^q([0,T];BV(\Omega))} \leq c,$$

whereas $\hat{p} = \min\{p, 2\}$. Consequently, $\mathcal{S}(c)$ is not empty and compact in the weak* topology.

A proof can be found in Dirks (2015).

Lemma 4.5. (Lower Semi-Continuity of $J_2(u, \boldsymbol{v})$)
The functional $J_2(u, \boldsymbol{v})$ as defined in (4.4) is weak* lower semi-continuous in $\mathcal{D} \subset \mathcal{X}$.

Proof. We use the same argumentation as in Lemma 4.2. Since again all summands of the functional $J_2(u, \boldsymbol{v})$ are functionals on Banach spaces, we obtain weak* lower semi-continuity for the complete functional. □

Finally, we achieve the existence of a minimal solution for the reformulated variant of Model 4.3:

Theorem 4.2. (Existence of Minimizers for the L^2-TV-TV-L^1 Model)
Let $1 < p, q$, $\hat{p} = \min\{p, 2\}$ and

$$J_2(u, \boldsymbol{v}) = \int_0^T \frac{1}{2}\|Hu - f\|_2^2 + \alpha|u(\cdot, t)|_{BV}^p + \beta|\boldsymbol{v}(\cdot, t)|_{BV}^q \, dt.$$

Furthermore, let

$$\|\boldsymbol{v}\|_\infty \leq c_v < \infty \quad \text{in } \Omega \times [0, T] \qquad \text{and} \qquad H\mathbf{1}_t \neq 0 \quad \forall t \in [0, T].$$

Then, there exists a minimizer in the set

$$\mathcal{S}(c) = \left\{(u, \boldsymbol{v}) \in \mathcal{D} \middle| J_2(u, \boldsymbol{v}) \leq c, \, \boldsymbol{v} \cdot \nabla u + \partial_t u = 0\right\}.$$

Proof. The functional $J_2(u, \boldsymbol{v})$ fulfills the requirements for the fundamental theorem of optimization. We obtain coercivity for $J_2(u, \boldsymbol{v})$ from Lemma 4.4 and lower semi-continuity from Lemma 4.5. The lower semi-continuity of the constraint is shown in Lemma 4.3. Consequently, there exists a minimizer for the reformulation of Model 4.3. □

4.4 Numerical Realization

In the course of this section we formulate a minimization algorithm for the previously described indirect tracking models. This algorithm in mainly based on Dirks (2015). The discretization of the occurring terms can be handled analog to Section 3.6. We refer to Dirks (2016b) for further details about the implementation including the choice of parameters for the models with nonlinear optical flow constraint. In both references an

indirect model with L^2 data fidelity for the reconstruction part is described. However, the respective implementation of the remaining terms besides the data fidelity for the reconstruction can be realized equivalently.

To solve the nonlinear optical flow constraint, we use the approximation, which we already introduced in Section 3.2.2, i.e.

$$\|u(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla u(\boldsymbol{v}^n)\|_1.$$

4.4.1 Alternating Minimization

In order to facilitate the numerical implementation of the models for indirect tracking, we use an alternating approach for the minimization, i.e. we alternately minimize an equation for the image u and an equation for the velocity field $\boldsymbol{v}$. For the models with linearized optical flow constraint the approach reads

$$\begin{aligned} u^{n+1} &= \arg\min_u \int_0^T \frac{1}{2}\|Hu - f\|_p^p + \alpha\|\nabla u\|_1 + \gamma\|\boldsymbol{v}\cdot\nabla u + \partial_t u\|_1\, dt, \\ \boldsymbol{v}^{n+1} &= \arg\min_{\boldsymbol{v}=(v_1,v_2)} \int_0^T \|\boldsymbol{v}\cdot\nabla u + \partial_t u\|_1 + \frac{\beta}{\gamma}\|\nabla \boldsymbol{v}\|_1\, dt \end{aligned}$$

with $p = 1$ for Model 4.2 and $p = 2$ for Model 4.3. For the nonlinear variants we obtain

$$\begin{aligned} u^{n+1} = \arg\min_{u,u_1} \int_0^T & \frac{1}{2}\|Hu - f\|_p^p + \alpha\|\nabla u\|_1 \\ & + \gamma\|u(\boldsymbol{v}^{n-1}) - u_1 + (\boldsymbol{v}^n - \boldsymbol{v}^{n-1})\nabla u(\boldsymbol{v}^{n-1})\|_1\, dt, \\ \boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}=(v_1,v_2)} \int_0^T & \|u^{n+1}(\boldsymbol{v}^n) - u_1^{n+1} + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla u^{n+1}(\boldsymbol{v}^n)\|_1 + \frac{\beta}{\gamma}\|\nabla \boldsymbol{v}\|_1\, dt \end{aligned}$$

with $p = 1$ for Model 4.4 and $p = 2$ for Model 4.5.

The equations for u and $\boldsymbol{v}$ can be solved separately with the Chambolle-Pock algorithm (Definition 2.24). For that purpose, we reformulate the single problems and use the notation

$$u^{n+1} = \arg\min_u F_u(K_u u) + G_u(u) \tag{4.5}$$

for the problem in u, respectively

$$\boldsymbol{v}^{n+1} = \arg\min_{\boldsymbol{v}} F_{\boldsymbol{v}}(K_{\boldsymbol{v}}\boldsymbol{v}) + G_{\boldsymbol{v}}(\boldsymbol{v}) \tag{4.6}$$

for the problem in $\boldsymbol{v}$. Each problem is separately minimized with the following iteration scheme:

$$\begin{aligned}
\tilde{y}^{k+1} &= y^k + \sigma K \bar{x}^k, \\
y^{k+1} &= \arg\min_{y} \left\{ \frac{\|y - \tilde{y}^{k+1}\|^2}{2\sigma} + F^*(y) \right\}, \\
\tilde{x}^{k+1} &= x^k - \tau K^* y^{k+1}, \\
x^{k+1} &= \arg\min_{x} \left\{ \frac{\|x - \tilde{x}^{k+1}\|^2}{2\tau} + G(x) \right\}, \\
\bar{x}^{k+1} &= x^{k+1} + \theta \left(x^{k+1} - x^k \right),
\end{aligned} \tag{4.7}$$

where x corresponds to the particular primal variable u respectively $\boldsymbol{v}$, and y to the dual variable.

The algorithms for solving the subproblems for u and $\boldsymbol{v}$ are explained in the proximate sections. We alternately solve those problems until the difference ϵ between two subsequent iterations is smaller than a certain threshold. We calculate ϵ by

$$\epsilon = \frac{|u^n - u^{n+1}| + |\boldsymbol{v}^n - \boldsymbol{v}^{n+1}|}{2|\Omega|},$$

where $|\cdot|$ denotes the sum of absolute values.

4.4.2 Subproblem Reconstruction

All terms of the minimization problem with respect to u contain an operator. The data fidelity contains the reconstruction operator H. The TV regularizer as well as the optical flow constraint contain the gradient operator ∇. In the linearized case the optical flow constraint additionally contains a temporal derivative. Consequently, we assign all of them to F_u and it holds

$$G_u(u) = 0$$

for this subproblem. The operator K_u is a composition of the previously described operators. For the linearized models this results in

$$K_u u = \begin{pmatrix} H \\ \nabla \\ (\partial_t, \nabla) \end{pmatrix} u.$$

The adjoint operator of K_u applied to the dual variable $\boldsymbol{p}$ is given by

$$K_u^* \boldsymbol{p} = H^* p_1 - \nabla \cdot p_2 - (\partial_t p_{3,1} + \nabla p_{3,2}).$$

For the nonlinear models, K_u applied to u is defined by

$$K_u u = \begin{pmatrix} H \\ \nabla \\ (1, \nabla) \end{pmatrix} u.$$

For the adjoint operator of K_u applied to $\boldsymbol{p}$ this leads to

$$K_u^* \boldsymbol{p} = H^* p_1 - \nabla \cdot p_2 - (p_{3,1} + \nabla p_{3,2}).$$

For both variants the iteration scheme (4.7) reads

$$\tilde{\boldsymbol{p}}^{k+1} = \boldsymbol{p}^k + \sigma_u K_u \bar{u}^k, \tag{4.8}$$

$$p_1^{k+1} = \arg\min_p \left\{ \frac{\|p_1 - \tilde{p}_1^{k+1}\|^2}{2\sigma_u} + F_{u1}^*(p_1) \right\}, \tag{4.9}$$

$$p_2^{k+1} = \arg\min_p \left\{ \frac{\|p_2 - \tilde{p}_2^{k+1}\|^2}{2\sigma_u} + F_{u2}^*(p_2) \right\}, \tag{4.10}$$

$$p_3^{k+1} = \arg\min_p \left\{ \frac{\|p_3 - \tilde{p}_3^{k+1}\|^2}{2\sigma_u} + F_{u3}^*(p_3) \right\}, \tag{4.11}$$

$$\tilde{u}^{k+1} = u^k - \tau_u K_u^* \boldsymbol{p}^{k+1}, \tag{4.12}$$

$$u^{k+1} = \arg\min_u \left\{ \frac{\|u - \tilde{u}^{k+1}\|^2}{2\tau_u} \right\}, \tag{4.13}$$

$$\bar{u}^{k+1} = u^{k+1} + \theta_u \left(u^{k+1} - u^k \right). \tag{4.14}$$

The exact form of the resolvent operator in (4.9) depends on the chosen norm for the data fidelity. For Model 4.2 and Model 4.4 similar calculations as for Lemma 3.18 yield

$$p_1^{k+1} = \frac{\tilde{p}_1^{k+1}}{1 + \sigma_u} - \frac{\sigma_u}{1 + \sigma_u} f. \tag{4.15}$$

For Model 4.3 and Model 4.5 we use Lemma 3.20 to achieve

$$p_1^{k+1} = \pi_1(\tilde{p}_1^{k+1}) := \min_{\tilde{p}_1^{k+1}} \left(1, \max_{\tilde{p}_1^{k+1}}(-1, \tilde{p}_1^{k+1}) \right). \tag{4.16}$$

The problem in (4.10) contains the adjoint of a TV norm. We apply Lemma 3.19 and obtain

$$p_2^{k+1} = \pi_\alpha(\tilde{p}_2^{k+1}) := \min_{\tilde{p}_2^{k+1}} \left(\alpha, \max_{\tilde{p}_2^{k+1}}(-\alpha, \tilde{p}_2^{k+1}) \right).$$

For the linearized models the resolvent operator in (4.11) can directly be obtained from Lemma 3.12, i.e.

$$p_3^{k+1} = \frac{\tilde{p_3}^{k+1}}{\sigma_u} + \begin{cases} -\frac{\gamma}{\sigma_u}\nabla u & \text{if } \rho\left(\frac{\tilde{p}_3^{k+1}}{\sigma_u}\right) < -\frac{\gamma}{\sigma_u}(\nabla u)^2 \\ \frac{\gamma}{\sigma_u}\nabla u & \text{if } \rho\left(\frac{\tilde{p}_3^{k+1}}{\sigma_u}\right) < \frac{\gamma}{\sigma_u}(\nabla u)^2 \\ -\frac{\rho\left(\frac{\tilde{p}_3^{k+1}}{\sigma_u}\right)}{\nabla u} & \text{if } |\rho\left(\frac{\tilde{p}_3^{k+1}}{\sigma_u}\right)| \leq \frac{\gamma}{\sigma_u}(\nabla u)^2 \end{cases},$$

with $\rho(p_3) := p_3 \cdot \nabla u + u_t$. For the nonlinear models we define ξ as

$$\xi = (1, \boldsymbol{v}^n - \boldsymbol{v}^{n-1}).$$

Hence, it is

$$\|\xi\|^2 = \sqrt{1 + (\boldsymbol{v}^n - \boldsymbol{v}^{n-1})^2}.$$

Additionally, we define

$$\rho(p_3) = \nabla u \cdot p_3 - u_1.$$

With this notation the resolvent operator in (4.11) is given by

$$p_3^{k+1} = \frac{\tilde{p_3}^{k+1}}{\sigma_u} + \begin{cases} -\frac{\gamma}{\sigma_u}\xi & \text{if } \rho\left(\frac{\tilde{p_3}^{k+1}}{\sigma_u}\right) < \frac{\gamma}{\sigma_u}\|\xi\|^2 \\ \frac{\gamma}{\sigma_u}\xi & \text{if } \rho\left(\frac{\tilde{p_3}^{k+1}}{\sigma_u}\right) < -\frac{\gamma}{\sigma_u}\|\xi\|^2 \\ -\frac{f\left(\frac{\tilde{p_3}^{k+1}}{\sigma_u}\right)}{\xi} & \text{if } \left|\rho\left(\frac{\tilde{p_3}^{k+1}}{\sigma_u}\right)\right| \leq \frac{\gamma}{\sigma_u}\|\xi\|^2 \end{cases}.$$

Due to the missing primal part G, (4.12) and (4.13) can simply be combined as

$$u^{k+1} = u^k - \tau_u K_u^* \boldsymbol{p}^{k+1}.$$

As a stopping criterion we use the primal-dual error (3.44), which we already introduced to minimize the models for direct tracking.

4.4.3 Subproblem Motion Estimation

The problem in u is equivalent to the problems studied in Chapter 3. Therefore, we assign the complete data fidelity to $G_{\boldsymbol{v}}$. According to Lemma 3.12, for the linearized models the resolvent operator for $G_{\boldsymbol{v}}$ is given by

$$v = \tilde{v} + \begin{cases} -\tau\nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u)^2 \\ \tau\nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u)^2 \\ -\frac{\rho(\tilde{v})}{\nabla u} & \text{if } |\rho(\tilde{v})| \leq \tau(\nabla u)^2 \end{cases}, \tag{4.17}$$

with $\rho(v) \coloneqq v \cdot \nabla u + u_t$. For the nonlinear models we useLemma 3.13 to achieve

$$v = \tilde{v} + \begin{cases} -\tau\nabla u & \text{if } \rho(\tilde{v}) < -\tau(\nabla u\,(x+\hat{v}))^2 \\ \tau\nabla u & \text{if } \rho(\tilde{v}) < \tau(\nabla u\,(x+\hat{v}))^2 \\ -\frac{\rho(\tilde{v})}{\nabla u(x+\tilde{v})} & \text{if } |\rho(\tilde{v}\,)| \leq \tau(\nabla u(x+\hat{v}))^2 \end{cases}, \tag{4.18}$$

with $\rho(v) \coloneqq u_0(x+\hat{v}) - u_1(x) + (v-\tilde{v}) \cdot \nabla u(x+\tilde{v})$.

The TV regularizer is assigned to $F_{\boldsymbol{v}}$. According to Lemma 3.19 the adjoint resolvent operator of $F_{\boldsymbol{v}}$ is given by

$$\boldsymbol{q} = \pi_{\frac{\beta}{\gamma}}(\tilde{\boldsymbol{q}}) \coloneqq \min_{\tilde{\boldsymbol{q}}}\left(\frac{\beta}{\gamma}, \max_{\tilde{\boldsymbol{q}}}(-\frac{\beta}{\gamma}, \tilde{\boldsymbol{q}})\right).$$

The operator $K_{\boldsymbol{v}}$ is given by the gradient ∇. The adjoint operator is hence the negative divergence $-\nabla\cdot$. Consequently, the iteration scheme for solving (4.6) reads

$$\begin{aligned}
\tilde{\boldsymbol{q}}^{k+1} &= \boldsymbol{q}^k + \sigma_{\boldsymbol{v}} K_{\boldsymbol{v}} \bar{\boldsymbol{v}}^k, \\
\boldsymbol{q}^{k+1} &= \pi_{\frac{\beta}{\gamma}}(\tilde{\boldsymbol{q}}^{k+1}), \\
\tilde{\boldsymbol{v}}^{k+1} &= \boldsymbol{v}^k - \tau_{\boldsymbol{v}} K_{\boldsymbol{v}}^* \boldsymbol{q}^{k+1}, \\
v^{k+1} &= \arg\min_{v}\left\{\frac{\|v - \tilde{v}^{k+1}\|^2}{2\tau} + G(x)\right\}, \\
\bar{\boldsymbol{v}}^{k+1} &= \boldsymbol{v}^{k+1} + \theta_{\boldsymbol{v}}\left(\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\right),
\end{aligned}$$

where the equation for v is either given by (4.17) or (4.18). Again, we follow the above described scheme until the primal-dual error (3.44) is below a threshold.

Altogether, an alternating application of the scheme derived in Section 4.4.2 and the scheme derived in this section solves the minimization of the indirect tracking model. A concrete application of this model including numerical results is shown in Chapter 6.

5 Application 1: Cell Tracking

We already analyzed the theoretical details of motion estimation in Chapter 3. In this chapter we discuss a real biological application for direct tracking, which is the transmigration of leukocytes. Leukocytes are a special kind of white blood cells. In this context, transmigration describes the process of cells actively leaving a blood vessel. In the following, we analyze the motion of the leukocytes during this process. As most real applications, the analysis of transmigrating leukocytes is attended by some specific challenges that require special attention. For this exemplary application we describe the relevance of motion estimation and explain the mathematical tasks that are necessary in order to obtain reasonable results.

The structure of the chapter is as follows: First, we explain the transmigration process in detail. Afterwards, we describe the data we use to investigate the transmigration process and the involved mathematical tasks. In Section 5.2 we explain a registration step, which we perform to balance global background motion in the data. The procedure of automatically finding the interesting regions is discussed in Section 5.3. In Section 5.4 we describe the motion estimation of single leukocytes, which contains shape transformations as well as intracellular density changes. Finally, we discuss our numerical results of motion estimation for transmigration of leukocytes.

5.1 Introduction

The process of leukocyte transmigration is highly involved in many autoimmune diseases. Leukocytes are part of the immune system of the body. Usually, their task is to defeat infectious diseases. However, if they start to act without control, at some point they also start to attack healthy cells. Two of the most common illnesses for which this

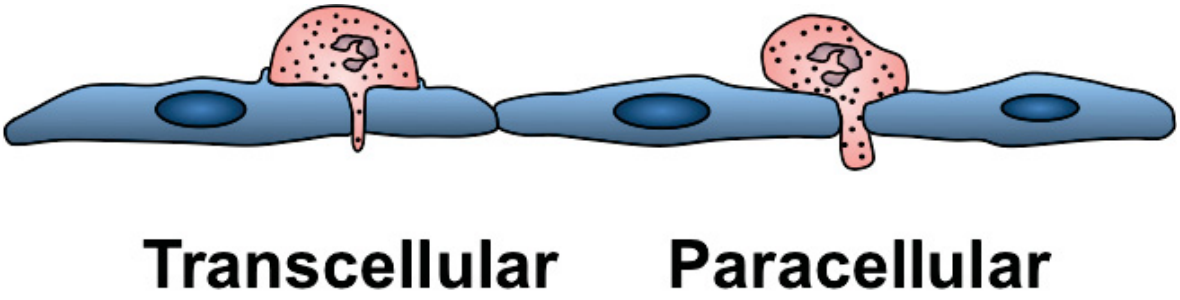

Figure 5.1: Different transmigration pathways. The blue cells represent endothelial cells; the red cells represent leukocytes during a transmigration process. Courtesy of Vestweber lab, MPI Münster.

occurs are multiple sclerosis and Crohn's disease. In therapy, it is possible to use immunosuppressive drugs, which reduce immune system activity. The consequence is that also real external intruders like viruses or parasites are not, or at least less, attacked. To develop more selective or more effective methods to help people who suffer from autoimmune diseases, it is necessary to understand the details about leukocyte transmigration. However, even if the underlying processes are known, there are still some aspects that raise questions.

In this chapter we discuss the fact that there are two different pathways taken by the leukocytes during the transmigration process. Up to now, it is not known why there are two possibilities, or whether it makes any difference for the leukocytes which one they use. Our aim is to get closer to answering these questions and thus learn more about the transmigration process.

5.1.1 Leukocyte Transmigration

Leukocytes are white blood cells that are usually located inside of blood vessels. During an injury in the body, a different amount of leukocytes moves out of the blood vessel and towards the damaged tissue and thus cause an inflammation (cf. Ley et al. (2007)). This process of leaving the blood vessel is called transmigration. The blood vessel itself is composed of different cell layers. The most inner one is called the endothelium. It consists of a single layer of endothelial cells that are connected at their cell boundaries. To leave the blood vessel during the transmigration process, the leukocytes have to cross the endothelium, which means that they have to find a way through the endothelial cells.

In general, there are two ways of crossing the endothelium. The first one is the transcellular pathway in which a leukocyte moves directly through an endothelial cell. For this purpose, it forms a hole into the endothelial cell, which gets closed again after

the leukocytes has passed. The second one is the paracellular pathway, which means that the leukocyte moves through the contacts between endothelial cells. It is either possible that it touches only two endothelial cells during this process, or it can as well move through a cross connection of three or more endothelial cells. Figure 5.1 shows a sketch of the two different transmigration pathways. On the left-hand side of the figure, a leukocyte (red) moves through an endothelial cell (blue) and thus moves transcellular. On the right-hand side, a leukocyte moves through the contact between endothelial cells, i.e. paracellular. Up to now there is only little known about the reasons for the occurrence of the different pathways. It seems that in certain regions of the body the leukocytes prefer one of them, which however does not mean, that they completely avoid the other one. It it even possible to block one of the ways, but the leukocytes do not transmigrate less in these cases. They just use the other transmigration pathway. For an analysis of the biological details of this topic cf. Vestweber (2007).

5.1.2 Data

The kind of data we want to analyze in this chapter is in vivo data of mice taken with a fluorescence microscope. It consists of three different channels: A PECAM-1 staining (red), which shows the endothelial cells, a VE-cadherin-eGFP staining (green), which shows the contacts between endothelial cells, and a Gr-1 staining (purple), which shows the leukocytes. The absence of PECAM-1 can be interpreted as possible transmigration pathway of leukocytes. Such openings occur if a leukocyte crosses the cell layer and meanwhile pushes the endothelial cells aside. However, in this staining it is not possible to decide whether the opening is inside a single endothelial cell, which would indicate a transcellular transmigration, or if it is located between two or more endothelial cells, which would indicate a paracellular transmigration. For this issue it is helpful to observe the channel that shows the contacts between the endothelial cells. If a leukocyte is migrating in a paracellular way, the contact between endothelial cells is interrupted at a certain point. Figure 5.2 shows close ups of a paracellular and a transcellular event, which are clearly identifiable by the combination of the PECAM-1 staining showing the endothelial cells and the VE-cadherin-eGFP staining showing the endothelial cell contacts. In the left images the gap in the PECAM-1 staining is obviously apart from every endothelial cell contact, which clearly indicates a transcellular transmigration. In the right images an endothelial cell contact is clearly interrupted by a gap, which reveals a paracellular transmigration.

Unfortunately, the staining of the contacts is not necessarily homogeneous even if there is no migration process at that time. Additionally, a gap in the staining caused by a

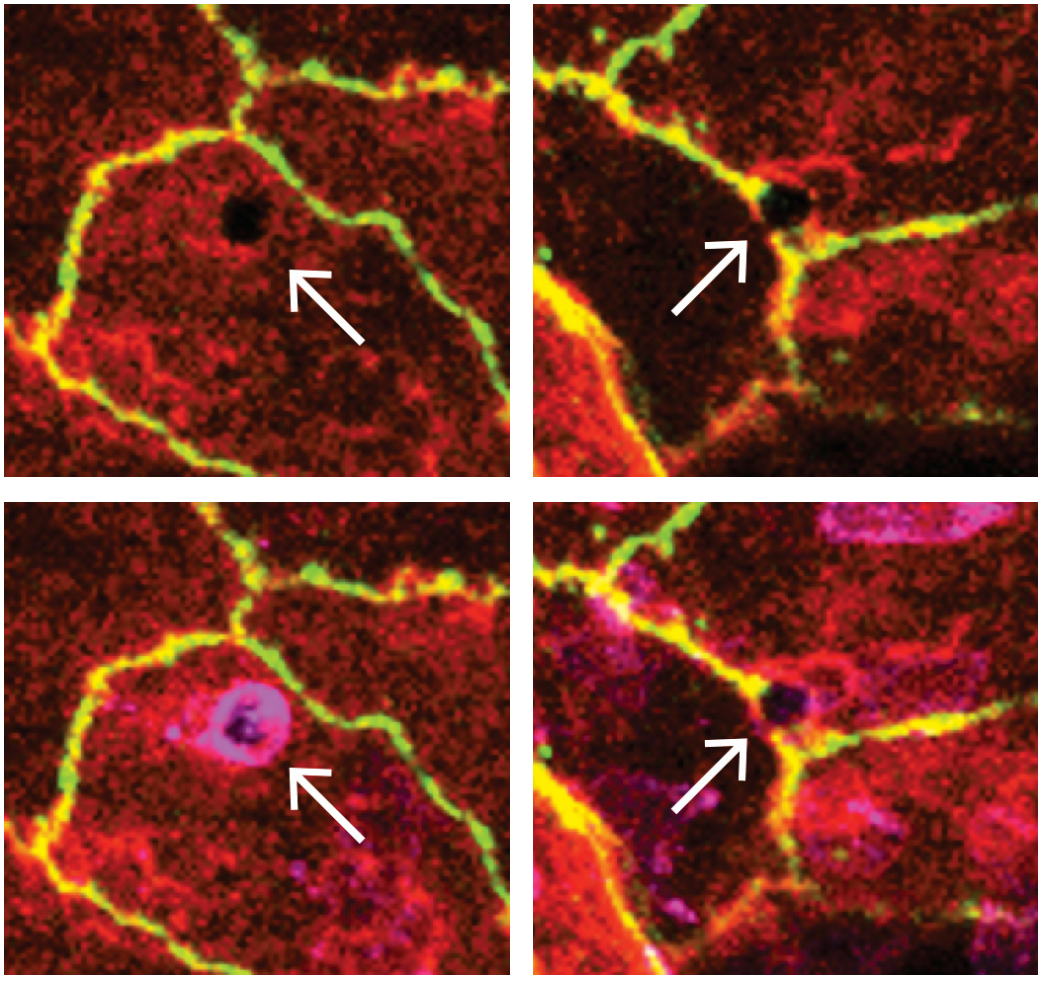

Figure 5.2: Crops of channels visualizing endothelial cells and endothelial cell contacts with transmigration events. Left: transcellular; right: paracellular, top: without Gr-1 staining, bottom: including Gr-1 staining. Data courtesy of Vestweber lab, MPI Münster.

leukocyte can be relatively small. Hence, it is not necessarily clear whether there is indeed a gap in the cell contact. In many cases a leukocyte is transmigrating very close to a cell boundary and pushes the boundary and its staining aside, but nevertheless it moves through the endothelial cell and not through the contact. Altogether, just considering the channel that shows the endothelial cell contacts is no sufficient criterion to automatically distinguish the different pathways. A possible solution to gain insight about the chosen pathway nevertheless, is to focus on the leukocytes themselves. Analyzing the transformation of their overall shape as well as their intracellular density, can potentially reveal significant differences, which allow to differentiate between the pathways.

While estimating the motion we have to face the fact that the leukocytes are moving quite fast compared to the temporal resolution of the data. A reason for this is that the recordings in z-direction are taken one after another. Every single layer takes a

certain time to be recorded, which depends on the resolution in x-y direction. To record a full three-dimensional stack, which combines to a single time step, it can take several minutes. Only if such a stack is complete, the recording for the subsequent time step can be started. During this time interval, the leukocytes naturally continue moving. It is possible that they cover distances of around 30 voxels in the x-y plane between consecutive time steps. There is also a small time difference between the recordings of the layers of a single time step. This is however usually small enough that it can be ignored for the later analysis.

The acquired data is three-dimensional with a resolution of $1024 \times 350 \times 39$ voxels in x-y-z-direction respectively. To be able to analyze the behavior of the leukocytes, the data is recorded during 75 consecutive time steps. Altogether, including the three channels, we are dealing with data of the size $1024 \times 350 \times 39 \times 75 \times 3$, which leads to approximately 3.1×10^9 voxels. Figure 5.3 shows a rendering of a three-dimensional microscopy data set of leukocytes in and around a blood vessel at a certain time step. A maximum intensity projection (MIP) of this data set can be seen in Figure 5.4.

5.1.3 Mathematical Tasks

Working with the previously described data, we have to deal with various mathematical challenges. The first difficulty is the fact that the data is taken in vivo. This means there is not only the local movement of the leukocytes, but also a global movement of the blood vessel and the whole background. This is e.g. caused by the cardiac activity or the respiration of the mouse. In order to be able to perform flow analysis, it is necessary to compensate for this global movement. Another challenge is the fact that it is not sufficient to analyze the different channels separately. To learn more about the leukocyte behavior, it is necessary to combine the information of the different channels. At least until there is more insight into the intracellular behavior, it is not possible to make assumptions about the pathway of a leukocyte by only regarding the Gr-1 staining. We need to incorporate its exact position related to the endothelial cell boundaries. On the other hand, only regarding the endothelial cells would also not be sufficient, since it is not always possible to decide if a hole in the cell interrupts a cell-cell contact, or if it is only very close to the boundary. Additionally, we usually have to deal with several transmigration processes at a time, which are located at different positions. In some cases there are around 15 events in the regarded part of the blood vessel at the same time. To better understand the whole process, we need to identify and analyze all events. Another obvious issue is the size of the data. We need to analyze data sets

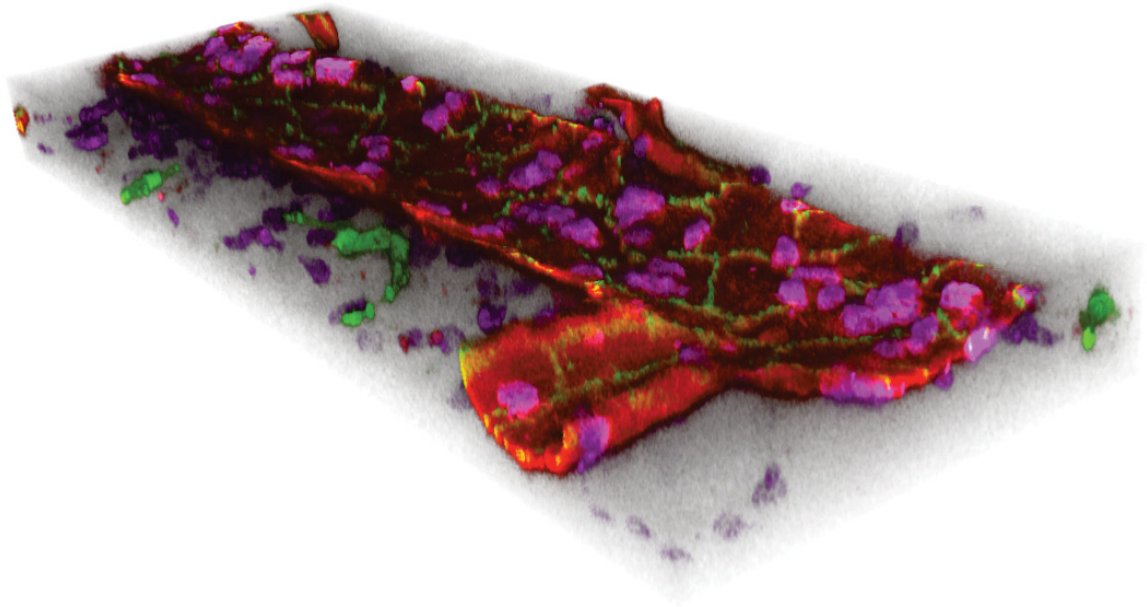

Figure 5.3: Three-dimensional microscopy data of leukocytes in and around a blood vessel. PECAM-1 staining (red): endothelial cells, VE-cadherin-eGFP staining (green): contacts between endothelial cells, Gr-1 staining (purple): leukocytes. Data courtesy of Vestweber lab, MPI Münster.

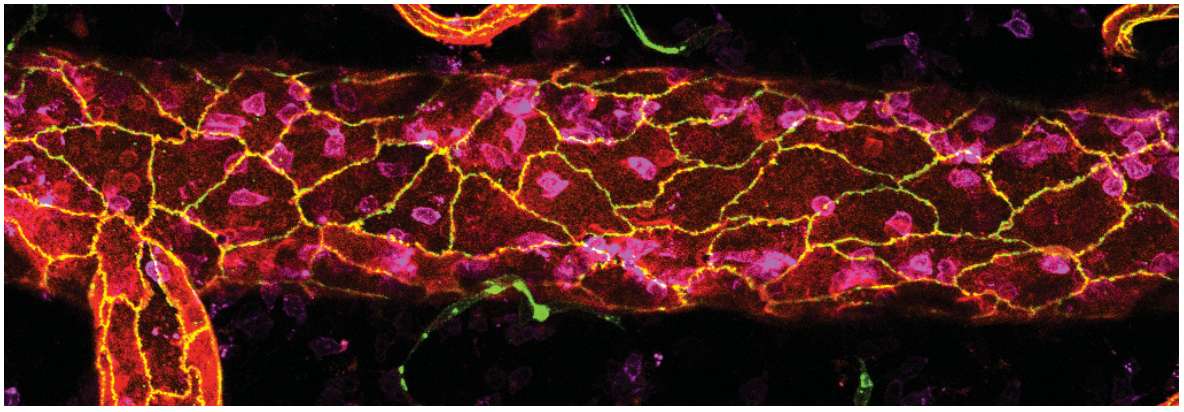

Figure 5.4: Maximum intensity projection of leukocyte data. PECAM-1 staining (red): endothelial cells, VE-cadherin-eGFP staining (green): contacts between endothelial cells, Gr-1 staining (purple): leukocytes. Data courtesy of Vestweber lab, MPI Münster.

with a size of more than three gigabyte. In order to be able to finish the analysis in a reasonable time, we have to reduce the amount of data and improve the efficiency of the used algorithms.

To handle all of the described difficulties, we start the data processing with an initial registration step. This serves to eliminate the global background motion and the blood vessel motion. Here, we have to pay attention to the fact that we will later need to combine the information of the different channels. Hence, we have to apply the exact same transformation to all of the different channels. Otherwise, the position of the leukocytes or the gap in the endothelial cell staining cannot be compared to the position of the cell-cell contacts anymore.

The next step is to find the regions of interest (ROIs) indicating a transmigration event. A transmigration process can only be located at positions at which there is a gap in the endothelial cell staining. We segment these gaps, which appear as black regions in the PECAM-1 staining. Later on, for every time step we will only have to analyze the ROIs and not the whole data set anymore. During this step we are also able to collect some useful information about the gaps like size, opening time and eccentricity. Similar to the gaps, we also segment the leukocytes in the Gr-1 staining. This way it is possible to connect both of the channels and check whether there is actually a leukocyte in the observed region. If this is not the case, there cannot be an ongoing transmigration event. Otherwise, observing such a gap with a leukocyte we also check the amount of consecutive time steps that the leukocyte remains in the gap. The segmentation of the leukocytes is also useful for a further aspect. With its help we are able to follow the paths of the leukocytes over a period of time. All of this information can give us hints for the different pathways.

However, there will usually be ROIs for which this kind of information is not sufficient. This is why we also focus on the leukocytes themselves to check whether we are able to find any specific differences for the two pathways in the intracellular behavior. We use the fact that we already know where the interesting regions are. We pick a single leukocyte that is indeed transmigrating and perform flow estimation. This means we analyze its respective density transformation. The idea is that we look for characteristic developments inside of the cell, e.g. a certain density due to a restricted gap size for one of the pathways. As mentioned before, the leukocytes move relatively fast compared to the temporal resolution of the data. That means we have to consider algorithms that handle large scale motion, in order to be able to analyze the local transformations. Figure 5.5 provides an overview of the workflow in this project.

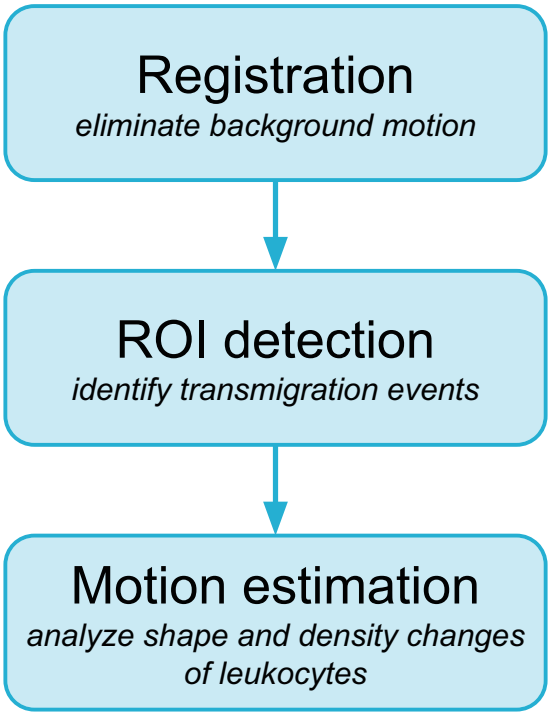

Figure 5.5: Workflow of the mathematical tasks.

5.2 Registration

Dealing with in vivo data is generally more challenging than dealing with in vitro, or even synthetic data. There are some additional aspects that have to be considered in the data analysis for in vivo data. One of them is that the studied animal is in general not absolutely immobile. There is at least some movement caused by cardiac and respiratory motion, also if the animal is narcotized. In contrast to the local movement of single cells, this kind of motion appears as a global movement of the whole crop in microscopic images. This is also the case in the blood vessel data sets we use. It can be observed that the blood vessel is shifted during the regarded time steps. Additionally, the blood vessel slightly changes its shape due to the blood circulation.

To compensate for this global movement, we register the data before starting to analyze smaller regions in detail. While doing that, we have to consider that we are dealing with three different channels. We are especially interested in the position of the leukocytes and of the gaps in the endothelial cells, both related to the contacts of the endothelial cells. Simply registering every channel independently of the other ones would most likely not end up with the exact same transformation for all channels. As a consequence, the location of a leukocyte could be slightly shifted regarding the position relative to the endothelial cell boundaries. This would make the whole analysis meaningless. To avoid

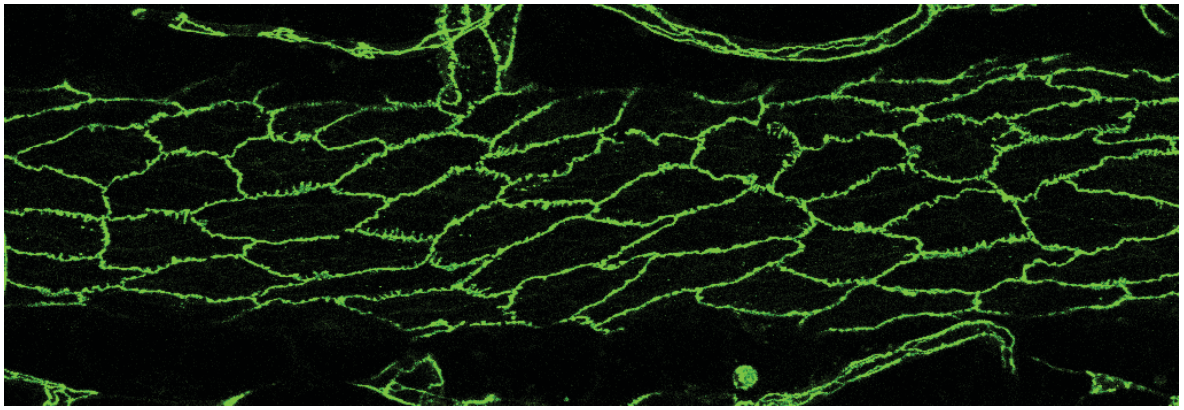

Figure 5.6: MIP of VE-cadherin-eGFP staining that visualizes the contacts between endothelial cells. Data courtesy of Vestweber lab, MPI Münster.

that we register a single channel, we save the transformation we used for this registration and also apply it to the other channels. The choice of the first channel to register is quite important, since not every channel is suited to estimate a proper transformation. In the channel with Gr-1 staining all of the single cells are moving independently, which makes it difficult to recover a global movement. The channel with PECAM-1 staining enables to find a global shifting, however a shape transformation of the endothelial cell is still difficult to detect, since the single cells are not easily distinguishable from each other. In contrast to that, the remaining channel with VE-cadherin-eGFP staining, which visualizes the contacts between the endothelial cells, has all of the missing features mentioned above. The motion in this channel is mainly caused by the global background motion. The independent motion of the endothelial cell contacts is very small. Additionally, the sharp edges in the staining allow for a simple recognition of exact positions in different time steps. Hence, it is possible to recover the global shifting, as well as the shape transformation of the endothelial cell. Thus, we use this one to correct all acquired data for the inherent background motion. Figure 5.6 shows the selected channel.

5.3 ROI Detection

Before starting a more detailed analysis, we have to reduce the amount of data. Otherwise, every used algorithm would need unnecessarily much time, and additionally most of the information we collect would be useless for us. Most of the leukocytes in the data set are not transmigrating at all, so it would not be helpful for us to analyze the behavior of each cell. We are only interested in analyzing the ROIs. If a transmigration event occurs, a leukocyte pushes parts of at least one endothelial cell aside. While that happens it is possible to see a gap in the PECAM-1 staining at this location. To find the

actual transmigration events, we only need to focus on regions around such gaps in the staining. However, the opposite argument does not work. Not every gap is necessarily the result of a transmigrating leukocyte. Nevertheless, the amount of data is drastically reduced by only focusing on the regions that contain a gap in the PECAM-1 staining.

In order to find the gaps we use simple thresholding to segment the dark areas in the PECAM-1 staining. If we denote the intensity in an image frame $\Omega \in \mathbb{R}^3$ at a certain time step by $u(x)$, with $x \in \Omega$, we use the functional

$$S(u) = \begin{cases} 0, \text{ if } u \leq \tau \\ 1, \text{ if } u \geq \tau \end{cases}$$

to distinguish between those parts that have higher intensity values than the threshold τ, and those that are smaller or equal to τ. This threshold is chosen manually. Since the background is dark as well, we only consider the segments that have a certain maximal size. This way we avoid to recognize the background as ROI. We identify the size of the areas with the help of the MATLAB-routine *regionprops* and then disregard those that are too large. However, since the staining of the endothelial cells is not continuous, there are many very small dark areas all over the blood vessel. To only consider those gaps with a reasonable size, we also neglect segments smaller than a certain minimal size. The maximal and minimal sizes are also chosen manually, since they are strongly dependent on the specific data set. However, to be sure not miss any transmigration events, we choose the maximal size rather large and the minimal size rather small. Gaps with a marginal size can be excluded manually at any time. The remaining areas are our ROIs. Figure 5.8 shows an overlay of the PECAM-1 staining and the result of the segmentation approach. *Regionprops* also enables us to easily collect information about the center position and the eccentricity of the gaps. Using the center position we are able to compare the gaps over a period of time. We start by giving a unique identification number (ID) to every segmented gap in the first time step. Then we regard the gaps in the second time step and check whether their center is close to the center of a gap in first time step. If that is the case we assume that the gap in the endothelial cell is the same and assign the same ID. If not we assume that the gap is not related to a previous one and assign a new ID. In an analog way, we proceed comparing gaps in consecutive time steps. Analyzing the appearances of the IDs, we obtain information about their opening time (OT).

We use the same thresholding algorithm as for the gaps in the endothelial cells to segment the leukocytes. Again, we only consider those segments that have a reasonable size. Afterwards, we combine the information about the leukocytes of the channel

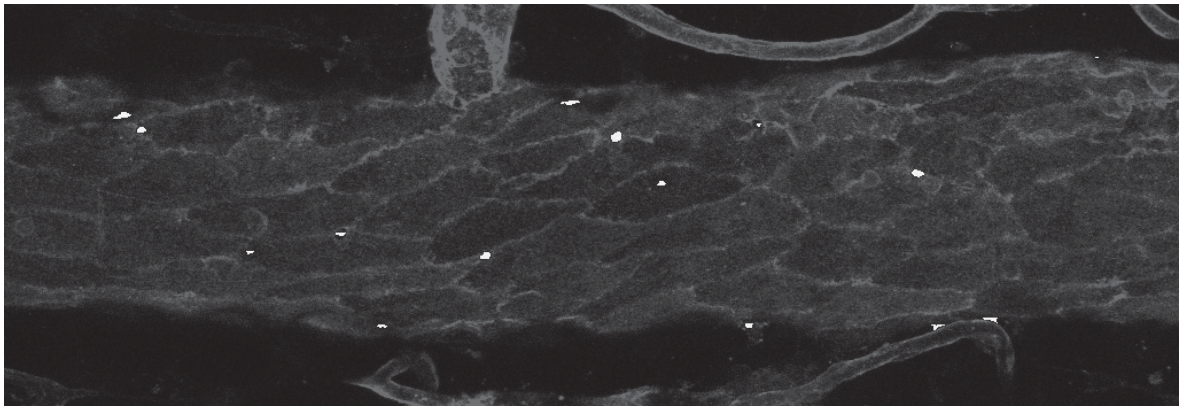

Figure 5.7: Overlay of MIP of a grayscale version of the PECAM-1 staining and the segmentation result highlighted in white. Data courtesy of Vestweber lab, MPI Münster.

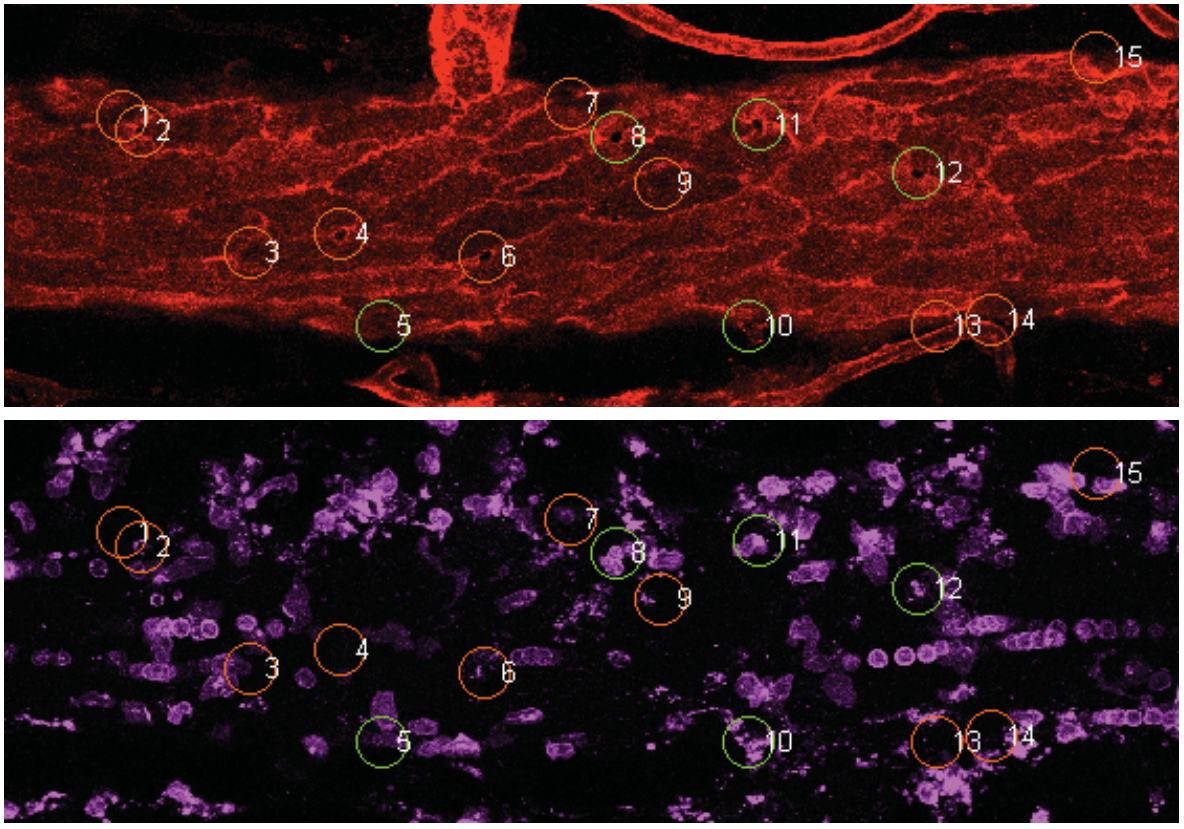

Figure 5.8: Top: MIP of PECAM-1 staining that visualizes the endothelial cells, bottom: MIP of Gr-1 staining that visualizes the leukocytes. Green marks highlight gaps that are currently occupied by a leukocyte; orange marks highlight gaps that are currently not occupied. Data courtesy of by Vestweber lab, MPI Münster.

ID	Position	Diameter	Eccentricity	OT	Occupancy
1	(103, 97)	8.21	0.95	1	
2	(120, 110)	6.38	0.77	6	2
3	(214, 216)	4.22	0.91	1	
4	(293, 200)	5.17	0.94	4	
5	(329, 280)	4.92	0.94	1	1
6	(420, 219)	7.40	0.77	3	
7	(494, 86)	7.31	0.98	2	
8	(534, 116)	9.03	0.64	3	3
9	(574, 157)	5.41	0.80	1	2
10	(650, 280)	5.53	0.77	2	2
11	(658, 106)	3.39	0.54	2	2
12	(797, 148)	7.98	0.83	4	3
13	(814, 281)	5.86	0.91	1	
14	(861, 275)	6.38	0.96	6	2
15	(953, 47)	1.95	0.04	1	

Table 5.1: Exemplary information about segmented gaps PECAM-1 staining and their occupancy by leukocytes.

with Gr-1 staining with the information about the endothelial cells of the channel with PECAM-1 staining and compare the particular positions of leukocytes and gaps. If their positions coincide at some time point, there is a good chance that a transmigration is happening at that position. Again, we regard this characteristic for all of the time steps to see how often a gap is occupied by leukocytes. Figure 5.8 shows the endothelial cells with highlighted gaps and the respective leukocyte occupancy at a single time step. Section 5.3 provides an overview of exemplary information collected during the segmentation step. Gaps without matching leukocytes at a certain time step might also be of interest, since it usually takes some time until the gap closes after a transmigration event. However, gaps that can never be matched with a leukocyte do not need to be analyzed any further.

5.4 Motion Estimation

To actually analyze the behavior of single leukocytes, we perform flow estimation as introduced in Chapter 3. In this application we are not only interested in tracking the path of the leukocytes, but also in analyzing their shape changes and the intracellular behavior. Regarding these characteristics we aim at identifying differences between the two transmigration pathways. In this section we describe the application of a

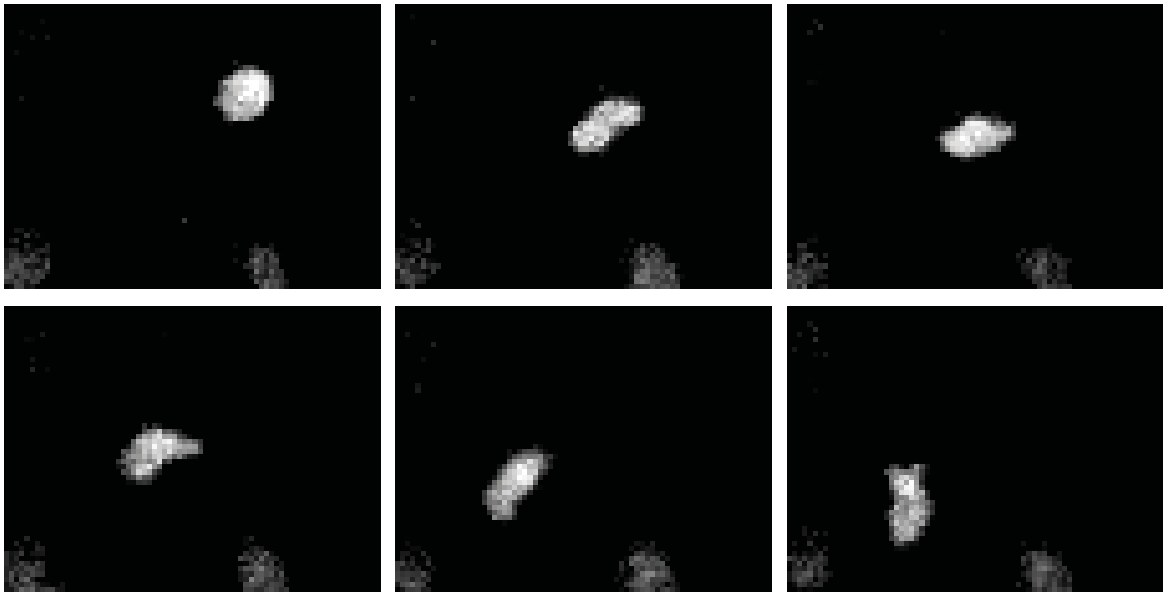

Figure 5.9: Transmigrating leukocyte in ROI with six consecutive time steps. Data courtesy of Vestweber lab, MPI Münster.

motion estimation model that enables a reasonable data analysis. The analysis itself is however more a biological than a mathematical task and is thus neglected in this context.

The motion estimation is based on the assumption that the intensity of the channel, which visualizes the leukocytes, is correlated to the density of the leukocytes. The brighter the image at a certain position is, the more staining is absorbed at that spot. Figure 5.9 shows a transmigrating leukocyte for which we want to perform motion analysis. Due to noise and possible spots caused by staining artifacts, fluorescence microscopy data typically tends to contain strong outliers concerning the intensity. For this reason we use an L^1 data fidelity, since it is known to be more robust than its L^2 equivalent. The leukocytes partially move about 30 pixels or more between two time steps. Therefore, we require a model that is able to handle such large scale motion. This specification excludes the linearized optical flow constraints. Hence, the data fidelity of our choice is the nonlinear optical flow constraint

$$\mathcal{D}_{NL1}(u_0, u_1, \boldsymbol{v}) = \|u_1 - u_0(\boldsymbol{v})\|_1.$$

For the minimization process, we approximate this data fidelity with the iterative functional, which we derived in Section 3.2.2, i.e.

$$\mathcal{D}_{approx}(u_0, u_1, \boldsymbol{v}) := \|u_0(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n)\|_1 + \frac{c}{2}\|\boldsymbol{v} - \boldsymbol{v}^n\|_2^2,$$

with $c > \sup_x \|H(u_0(x))\|$ and H being the Hessian matrix.

Since we have no information about the ground truth flow, we need to use a priori information from biological knowledge about the behavior of the leukocytes in order to choose a proper regularizer. An important feature in this context is the fact that, at least after the registration step, the background contains no, or only very small movements. In contrast to that, the movement of the entire leukocyte is very fast. For the velocity field this means that we assume to have strong edges at the locations of the cell boundaries, since this is where the fast movement hits the slow or zero movement. For this reason a TV-based regularizer is assumed to be a good choice. However, a leukocyte is no constant object. Its intracellular texture is viscous and thus changes during the microscopic recording. Due to the viscosity, the density changes inside the leukocyte are supposed to be relatively smooth and do not usually contain strong transitions. Therefore, the regularizer should also be able to detect such smooth transitions. In Section 3.3.2 we developed regularizers, which combine both capabilities. We specified two variants for the extended regularizer, which we both want to use for estimating the motion of the leukocytes. The first variant is the TV/L^2 regularizer

$$\mathcal{R}_{TV/L^2}(\boldsymbol{v}, w) = \|\nabla v - w\|_1 + \frac{\alpha_2}{2\alpha_1} \|w\|_2^2 .$$

With this choice the minimization functional reads

$$\begin{aligned} J_{TV/L^2}(u_0, u_1, \boldsymbol{v}) = \min_{\boldsymbol{v}} \Big\{ & \|u_0(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n)\|_1 \\ & + \min_{w} \Big\{ \alpha_1 \|\nabla v - w\|_1 + \frac{\alpha_2}{2} \|w\|_2^2 \Big\} \Big\} . \end{aligned}$$

The second variant is the TV/TV regularizer

$$\mathcal{R}_{TV/TV}(\boldsymbol{v}, w) = \|\nabla v - w\|_1 + \frac{\alpha_2}{2\alpha_1} \|\nabla w\|_1,$$

with the corresponding minimization functional

$$\begin{aligned} J_{TV/TV}(u_0, u_1, \boldsymbol{v}) = \min_{\boldsymbol{v}} \Big\{ & \|u_0(\boldsymbol{v}^n) - u_1 + (\boldsymbol{v} - \boldsymbol{v}^n)\nabla u_0(\boldsymbol{v}^n)\|_1 + \frac{c}{2}\|\boldsymbol{v} - \boldsymbol{v}^n\|_2^2 \\ & + \min_{w} \Big\{ \alpha_1 \|\nabla v - w\|_1 + \frac{\alpha_2}{2} \|\nabla w\|_1 \Big\} \Big\} . \end{aligned}$$

In contrast to the synthetic data sets in Section 3.7, for this real data set it is not possible to choose the best regularization parameters α_1 and α_2 based on the minimal errors AEE and AE. The only possibility to choose them properly is a visual evaluation of the results.

Figure 5.10 shows the motion estimation results for the model with TV/L^2 regulariza-

Figure 5.10: Result of TV/L^2 model for transmigrating leukocyte at 6 consecutive time steps.

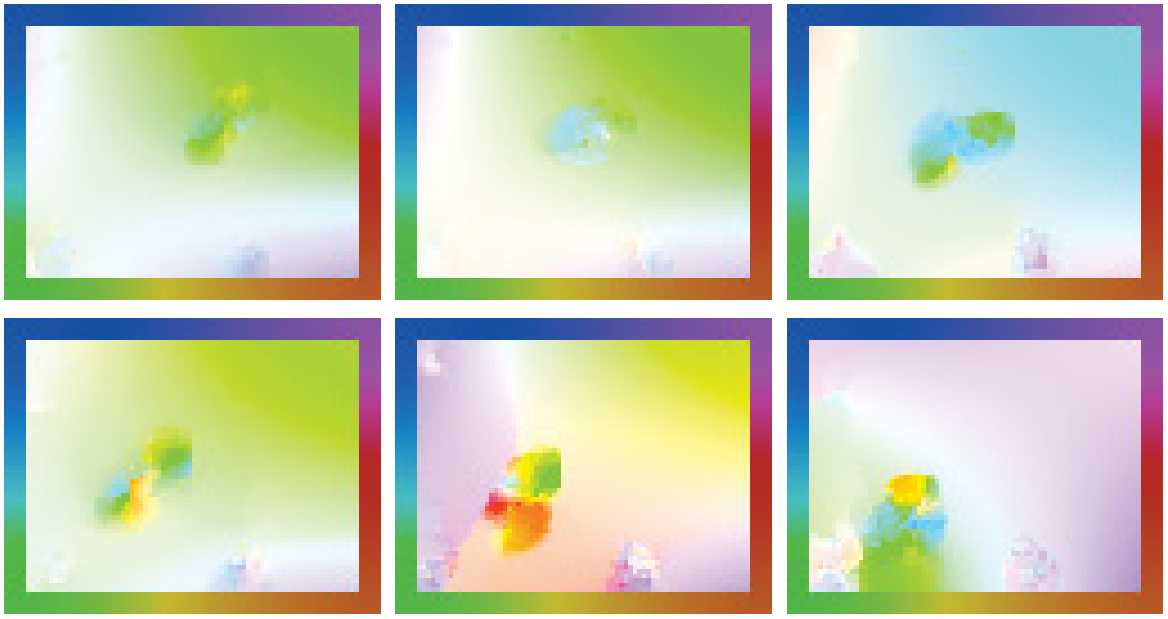

Figure 5.11: Result of TV/TV model for transmigrating leukocyte at 6 consecutive time steps.

tion. The depicted ROI corresponds to Figure 5.9. In these results the boundaries of the leukocyte are clearly visible as edges in the velocity fields. However, there is also motion detected in the background. This is a consequence of an almost constant area. Without any structure in the data, the optical flow constraint can also be fulfilled for incorrect movement due to the aperture problem. Nevertheless, the actual movement of the leukocyte is well identifiable. In contrast to the clearly visible edges at the cell boundaries, the flow of the intracellular texture of the leukocyte appears smooth without any sharp edges. Even though there are changing directions in this area, there is no occurring staircasing effect as we know it from standard TV regularization. The intracellular flow indeed shows a realistic behavior. Consequently, the model performs as intended and therefore allows an interpretation from a biological point of view.

Figure 5.11 illustrates the results for the same scenery calculated with the TV/TV regularization. A noticeable characteristic of these results is the active background of the ROI. The flow fields from the TV/L^2 results do only contain consistent motion in the background, at least in space. In contrast to that, in the TV/TV results the background motion is changing in time and in space. Even though the estimated motion of the cell itself is very similar for both models, the edges at the boundaries appear to be sharper for the TV/L^2 model due to the clearer contrast to the background.

6

Application 2:

Dynamic X-ray Tomography

In this chapter we show a concrete application for the indirect tracking model, which we derived in Chapter 4. The idea for this application is based on data from an X-ray lab owned by the University of Helsinki. The X-ray scanner of this lab consists of only a single pair of source and detector. To reconstruct dynamic X-ray data it is usually necessary to be able to reconstruct every single time step independently. This requires an amount of about 15 pairs of source and detector that scan each time step (cf. Bian et al. (2010)). The aim of this chapter is to enable dynamic X-ray tomography for scanners with the more limited amount of only one or two X-ray sources by using the indirect tracking model. The results in this section originate from a cooperation with Samuli Siltanen, Tapio Helin and Andreas Hauptmann from the University of Helsinki, as well as Martin Burger and Hendrik Dirks from the University of Münster.

We start with a short introduction to the general topic of dynamic X-ray tomography in Section 6.1. We continue by illustrating the mathematical details in Section 6.2. Since the reconstruction of X-ray tomography requires the usage of a Radon transform, we especially focus on this issue. We describe the details for the standard and for a time-dependent version and use the latter as reconstruction operator for the indirect tracking models from Chapter 4. Subsequently, we explain the experimental settings for the following numerical approaches. The settings include an initialization for the first and the last time step, the usage of multiple angles per time step, and a randomized

distribution of subsequent angles to increase the spatial area that is used for subsequent time steps. Finally, we evaluate the different strategies by showing numerical results for two different data sets. These contain a synthetic data set to test the performance of the mathematical model, as well as a real measured data set.

6.1 Introduction

The research on tomographic imaging was initiated in the seminal work of Johann Radon on the determination of functions from their integrals along certain manifolds, which he published already about 100 years ago (Radon (1917)). These findings about the meaning of the line integrals are of great importance for a wide range of applications, cf. Cormack (1963) and Cormack (1964). The most popular application is computerized tomography. The basic principle behind this is to measure the line-integrals of the density of an object from a distinct amount of angles and combine the information gained from all angles in a sinogram. With the help of a Radon transform, which we describe more precisely in Section 6.2.2, this sinogram is used to reconstruct the given object. Even though the mathematics behind computerized tomography have already been extensively studied, e.g. in Natterer (1986), there are still some meaningful open problems concerning this issue.

A topic that received more and more attention in the past few years is dynamic tomography. This indicates the time-dependent measurement of moving objects. In a medical context, a goal is e.g. to analyze organs that are non-stationary throughout the scan. Whereas the reconstruction of dynamic data is already possible in case of full-angle measurements at each time step, a limited-angle reconstruction is still an open problem. The limitation of angles is however a common topic especially in medical applications. Besides the fact that it is quite costly to place sources and detectors all around a body in order to be able to scan every angle at every time step, the main drawback of full-angle measurements is the ionizing effect of the X-rays. In living organisms this can cause severe damages by e.g. drastically increasing the cancer risk. Scanning an object from every angle at every considered time step induces an extremely high radiation dose. Reducing the amount of necessary angles to be measured per time step also reduces the radiation dose and is therefore of special interest. However, this also reduces the amount of information obtained by a single scan and hence has a trade-off.

In this chapter we use a setting in which the number of measured angles is being limited drastically, i.e. we allow for only one or two measurements per time step. While doing this, a problem arises due to the missing information at the single time steps. As

already mentioned previously, the reconstruction of X-ray tomography data relies on the combined information from the measurements of all given angles. In our setting there is only information gained by the measurements of one or two angles for the single time steps. Hence, reconstructing them independently from each other cannot produce reasonable results. The principle of our approach is to fill this lack of information with the help of the motion field. Thus, we are able to connect the different time steps with each other and use the information of all measured angles.

The indirect model we derived in Chapter 4 is able to fulfill the just described task. It uses the motion field to improve the reconstruction of the given data. The motion field, which is not initially given in our setting, is estimated based on the previously calculated reconstruction. For the following study we use Model 4.4 and Model 4.5, which we specified in Section 4.3.1. The first data fidelity $\mathcal{D}_{rec}(u, f)$, which solves the reconstruction problem, contains a time-dependent Radon transform. We will further discuss this issue in the following section. The remaining terms as well as the notation can simply be adopted from Section 4.3.1.

6.2 X-Ray Tomography

In this section we explain the mathematical details of X-ray tomography. To get an idea of the basic principle, we start with a short introduction to stationary X-ray tomography. In Section 6.2.2 we focus on the specialties of dynamic X-ray tomography, which requires a time-dependent formulation of the Radon transform. Subsequently, we incorporate the time-dependent variant as reconstruction operator in the models for indirect tracking from Section 4.3.1 and prove the requirements for the existence of a minimal solution.

6.2.1 Stationary Radon transform

The goal of common non-dynamic X-ray tomography is to reconstruct a two dimensional image slice u based on several projections of the volume of a given object from different angles. The projections occur from X-rays that are sent out from an X-ray source. A detector, which is located at the opposite side of the source, receives measurements that are given as integrals along lines through the object. The Radon transform is the functional that describes the connection between the given object and the measured data. The sinogram, i.e. the collection of the measurements from all given angles, serves as a basis for the reconstruction of the image slice by applying the inverse Radon transform.

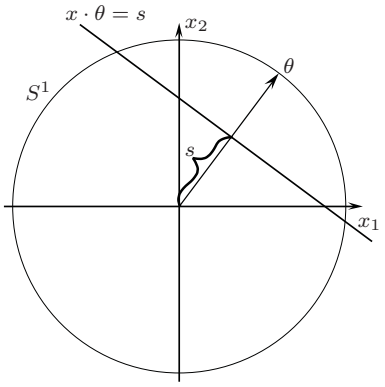

Figure 6.1: Illustration of the Radon transform in $\mathbb{R}^2$ taken from HAUPTMANN (2012).

The problem we aim to solve for the reconstruction is

$$\mathcal{R}u = m,$$

where $\mathcal{R}$ is the Radon transform, u is the desired image slice, and m is the sinogram. The image slice u depicts the non-negative absorption of photons an a domain $\Omega \subset \mathbb{R}^2$. Hence, it is $u : \Omega \to \mathbb{R}_+$. The lines through the object are dependent on an angle $\theta \in [0, \pi)$ and a signed distance from the origin, which we denote by $s \in \mathbb{R}$. They are thus given by

$$L(\theta, s) = \left\{ (x_1, x_2) \in \mathbb{R}^2 \,\middle|\, x_1 \cos(\theta) + x_2 \sin(\theta) = s \right\}.$$

The Radon transform, which describes the line integrals, is defined by

$$(\mathcal{R}_I u)(\theta, s) := \left(\int_{L(\theta,s)} u(x)\, dx \right)_{(\theta,s)\in I},$$

where I indicates the set of given measurements. Consequently, the Radon transform is a mapping onto a parametrization of the infinite unit cylinder in $\mathbb{R}^3$, i.e.

$$Z^2 = \left\{ (\theta, s) \,\middle|\, \theta \in S^1, s \in \mathbb{R} \right\},$$

where S^1 is the unit circle in $\mathbb{R}^2$. An illustration of the Radon transform is shown in Figure 6.1.

The Radon transform is a bounded and linear operator. For the static case it is possible to uniquely determine the attenuation coefficient for every point $x \in \Omega$ and therefore the requested image slice u, see RADON (1917). This however requires the incorporation of

the full sinogram, which includes every possible angle. For further details and properties of the Radon transform, cf. NATTERER AND WÜBBELING (2001).

6.2.2 Time-dependent Radon Transform

In contrast to stationary X-ray tomography, in the dynamic case the image slice u is not only dependent on the spatial position $x \in \Omega \subset \mathbb{R}^2$, but also on the current time step $t \in [0, T]$, with $T > 0$. For the following analysis we assume that the chosen time interval $[0, T]$ is fixed and bounded. For the single time steps we are not able to access a full sinogram. Due to our assumptions, there is only a limited amount of angles available at every step. The exact selection of angles is different for every time step. Together with the fact that the measurements differ in case of a moving scenery, this also leads to a time dependency of the sinogram, i.e. $m(\theta, s, t)$. Consequently, we are dealing with a time-dependent Radon transform

$$\mathcal{A} : L^2(\Omega \times [0, T]) \to L^2(\cup_{t\in[0,T]} I(t) \times \{t\}),$$

in order to solve the inverse problem

$$\mathcal{A}u(x, t) = m(\theta, s, t). \tag{6.1}$$

We define the adapted Radon transform $\mathcal{A}$ by

$$(\mathcal{A}u)(\theta, s, t) := (\mathcal{R}_{I(t)}u(x, t))(\theta, s) = \int_{L(\theta,s)} u(x, t)\, dx, \tag{6.2}$$

where $\mathcal{R}_{I(t)}$ is the stationary Radon transform introduced in the previous section.

The time-dependent Radon transform is well-defined:

Proposition 6.1. (Well-Definedness of the Time-Dependent Radon Transform)
Let $[0, T]$ be a fixed and bounded time interval. Then for $p \geq 1$ the time-dependent Radon transform $\mathcal{A} : L^p(\Omega \times [0, T]) \to L^p(\cup_{t\in[0,T]} I(t) \times \{t\})$, as defined in (6.2), is well-defined.

Proof. We start with the case $p = 1$. For the measuring angle θ the signed distance s

from the origin and the width p of the measurement, the line integral $\mathcal{R}_{I(t)}$ is given by

$$\begin{aligned}\mathcal{R}_{I(t)}u(\theta,s,p,t) &= \int_0^T \int_{\mathbb{R}} \left| \int_{\mathbb{R}} u(s\theta(t)+p\theta^{\perp}(t))\,dp \right| ds\,dt \\ &\leq \int_0^T \int_{\mathbb{R}} \int_{\mathbb{R}} \left| u(s\theta(t)+p\theta^{\perp}(t)) \right| dp\,ds\,dt.\end{aligned} \tag{6.3}$$

For a fixed $t \in [0,T]$ we perform a change of variables by writing

$$x := s\theta(t) + p\theta^{\perp}(t).$$

Consequently, it holds

$$dx = ds\theta(t) + dp\theta^{\perp}(t).$$

Inserting this into (6.3) yields

$$\begin{aligned}\mathcal{R}_{I(t)}u(x,t) &\leq \int_0^T \int_{\Omega(t)} |u(x)|\,dx\,dt \\ &\leq \int_0^T \int_{\Omega} |u(x)|\,dx\,dt \\ &= \mathcal{A}u(x,t),\end{aligned}$$

where $\Omega(t) \subset \Omega$ denotes the domain of the single measurement at time step t. Therefore, the time-dependent Radon transform is a well-defined operator

$$\mathcal{A} : L^1(\Omega \times [0,T]) \to L^1(\cup_{t\in[0,T]} I(t) \times \{t\}).$$

For $p > 1$ the calculations can be performed equivalently. □

Moreover, we are able to show that the time-dependent Radon transform is continuous:

Proposition 6.2. (Continuity of the Time-Dependent Radon Transform)
Let $[0,T]$ be a fixed and bounded time interval. Then the time-dependent Radon transform $\mathcal{A} : L^2(\Omega \times [0,T]) \to L^2(\cup_{t\in[0,T]} I(t) \times \{t\})$ as defined in (6.2) is continuous.

Proof. The time-dependent Radon transform consists of consecutive applications of the stationary Radon transform at several time steps. Since the interval $[0,T]$ is bounded and hence finite, the continuity of the time-dependent Radon transform is a direct consequence, of the continuity of the stationary Radon transform. □

For the following numerical realization, we aim at minimizing the specific models including the time-dependent Radon transform with the primal-dual algorithm, which

we presented in Section 4.4.1. For this purpose, we calculate the adjoint $\mathcal{A}^*$ of $\mathcal{A}$:

Proposition 6.3. (Adjoint of the Time-Dependent Radon Transform)
Let $[0,T]$ be a fixed and bounded time interval, $\mathcal{A} : L^2(\Omega\times[0,T]) \to L^2(\cup_{t\in[0,T]} I(t)\times\{t\})$ the time-dependent Radon transform, and $m \in L^2(\Omega \times [0,T])$. Then the L^2 adjoint of $\mathcal{A}$ is given by

$$(\mathcal{A}^* m)(x,t) = ({\mathcal{R}^*}_{I(t)} m_t)(x) = \int_{I(t)} m(\theta, x\cdot\theta) d\theta,$$

where ${\mathcal{R}^*}_{I(t)}$ is the L^2 adjoint of the stationary Radon transform.

Proof. The proof follows directly from the properties of the stationary Radon transform and the following calculations:

$$\begin{aligned}(\mathcal{A}u, m)_{L^2(Z^2\times\mathcal{T})} &= \int_0^T \int_{I(t)} \int_{\mathbb{R}} (\mathcal{R}u_t)(\theta,s) m_t(\theta,s)\; ds d\theta dt \\ &= \int_0^T \left(\mathcal{R}_{I(t)} u_t, m_t\right)_{L^2(Z^2)} dt \\ &= \int_0^T \left(u_t, R^*_{I(t)} m_t\right)_{L^2(\Omega)} dt = (u, \mathcal{A}^* m)_{L^2(\Omega\times\mathcal{T})}\end{aligned}$$

□

6.2.3 Joint Tracking and Reconstruction

We aim to use the time-dependent Radon transform in combination with a motion estimation task. Therefore, we incorporate the operator into the data fidelity of the specified models for indirect tracking from Section 4.3.1. Since in realistic applications we usually have to deal with large scale motion, we only consider the models with the nonlinear optical flow constraint in this chapter. For Model 4.4 we obtain

$$\arg\min_{u,\mathbf{v}} \int_0^T \|\mathcal{A}u - m\|_1\, dt + \alpha\|\nabla u\|_1 + \beta\|\nabla \mathbf{v}\|_1 + \gamma\|u_1 - u(\boldsymbol{v})\|_1\, dt.$$

Equivalently, for Model 4.5 we achieve

$$\arg\min_{u,\mathbf{v}} \int_0^T \frac{1}{2}\|\mathcal{A}u - m\|_2^2\, dt + \alpha\|\nabla u\|_1 + \beta\|\nabla \mathbf{v}\|_1 + \gamma\|u_1 - u(\boldsymbol{v})\|_1\, dt.$$

6.3 Experimental Settings

We aim at minimizing the amount of necessary measurements per time step for the dynamic X-ray tomography as far as possible. In an ideal case we only need a single measurement from a distinct angle for every step. This way the X-ray dose is as small as possible. However, a pure application of the indirect model to data including only a single angle per time step cannot lead to sufficiently reasonable results, since the lack of information is still too large. For this reason, we experiment with different scenarios that incorporate varying kinds of additional information. In this section we introduce different settings, which we will later use to analyze the numerical performance of the indirect tracking model applied to dynamic X-ray tomography.

6.3.1 First and Last Angle Initialization

In the first setting we assume to have extensive information about the scanned object during the first and during the last time step. From a medical point of view, this assumption makes sense in case of motion, which only starts at a specific time and ends at some later point. An example for this is the inflow of a contrast agent into the vessel system. Before adding the contrast agent, the visible scenery is nearly static besides some smaller movements caused by e.g. respiration, which can be subtracted in a post processing step. During the inflow of the contrast agent, there is a clearly visible motion, which is supposed to be analyzed. After some time, the contrast agent is available in every area of the vessel system, such that the scenery is again widely static. In this setting, it is possible to perform a full scan from all possible angles for the static situations, and nevertheless minimize the amount of necessary measurements during the remaining time steps. Since the different time steps are connected via the calculated flow field, the full scans act as a kind of initialization for all time steps. This way they can possibly equal the necessary information that enables a proper reconstruction for every single time step. For the numerical approaches, we consider measurements from 60 equally distributed angles from zero to 179 degree for the first and for the last time step. Between those scans we expect to have 28 further time steps at which we measure one consecutive angle per step. The different angles are again equally distributed in a range from zero to 179 degree. Hence, the available measurements $I(t)$ are given by

$$I(t) = \begin{cases} \left(\begin{pmatrix} 0 & \frac{1}{60}\pi & \frac{2}{60}\pi & \cdots & \frac{60}{60}\pi \end{pmatrix}^T, s \right), & \text{for } t = 1 \\ \left(\frac{t-1}{30}\pi, s \right), & \text{for } t \in \{2, \cdots, 29\} \\ \left(\begin{pmatrix} 0 & \frac{1}{60}\pi & \frac{2}{60}\pi & \cdots & \frac{60}{60}\pi \end{pmatrix}^T, s \right), & \text{for } t = 30 \end{cases}.$$

6.3.2 Multiple Angles

As a second possibility to overcome the lack of information is to consider more than only a single angle for every time step. For real measurements, this naturally requires multiple X-ray sources and detectors. To keep the amount of necessary sources and detectors as small as possible, we only consider two simultaneous measurements per time step. The problems caused by a single consecutive angle per time step arise due to the lack of information about the scanned object in the direction orthogonal to the projection. Also measurements of the surrounding angles, which usually show a very similar scenery if the movement is assumed to be small enough, cannot give more clarification of the orthogonal direction. Consequently, it takes multiple time steps until there is sufficient information available about the situation in the orthogonal direction to a particular projection. In case that there is motion taking place in the orthogonal direction during the previous time steps, this cannot be tracked properly. The most efficient way to place two X-ray sources and detectors is thus an orthogonal arrangement. This way the directions with the biggest lack of information with respect to one of the pairs of source and detector are scanned by the additional pair. For the concrete experiments we again consider 30 time steps. The first source-detector pair measures one consecutive angle at every time step, whereas the angles θ are equally distributed between zero and 179 degree. The other source-detector pair measures the angles at $\theta + 90$ degree. Consequently, we obtain

$$I(t) = \left(\begin{pmatrix} \frac{t-1}{30}\pi & \frac{t+14}{30}\pi \end{pmatrix}^T, s \right).$$

6.3.3 Random Angles

A different attempt to overcome the lack of information, especially in approximately orthogonal directions to the current projection, is to depart from the idea of measuring adjacent angles during consecutive time steps. For this approach, we only measure a single angle per time step, but in a random arrangement. Thus, for this setting the measurements $I(t)$ are given by

$$I(t) = \left(\frac{r(t) - 1}{60}\pi, s \right),$$

with $r(t) \in \{0, \cdots, 60\}$. We are aware that the specific choice of the angles is essential for the quality of the results. The closer two angles measured at consecutive time steps are, the less information about movements in the orthogonal direction is available. Nevertheless, we simply use the MATLAB function *rand*, to pick the respective angles.

Time Step	30 Ang. Pinball	30 Ang. Stones	15 Ang. Stones	8 Ang. Stones
1	109	72	126	129
2	48	9		
3	146	129	3	
4	44	153		
5	133	51	48	177
6	105	84		
7	171	30	6	
8	37	102		
9	79	126	15	75
10	37	39		
11	135	87	147	
12	38	24		
13	111	3	123	39
14	152	129		
15	143	108	57	
16	43	174		
17	93	33	171	129
18	59	144		
19	18	81	6	
20	172	27		
21	120	144	78	159
22	89	45		
23	35	141	66	
24	28	144		
25	91	0	135	114
26	156	51		
27	3	153	141	
28	73	177		
29	21	96	33	84
30	164	51		

Table 6.1: List of randomly measured angles for given time steps.

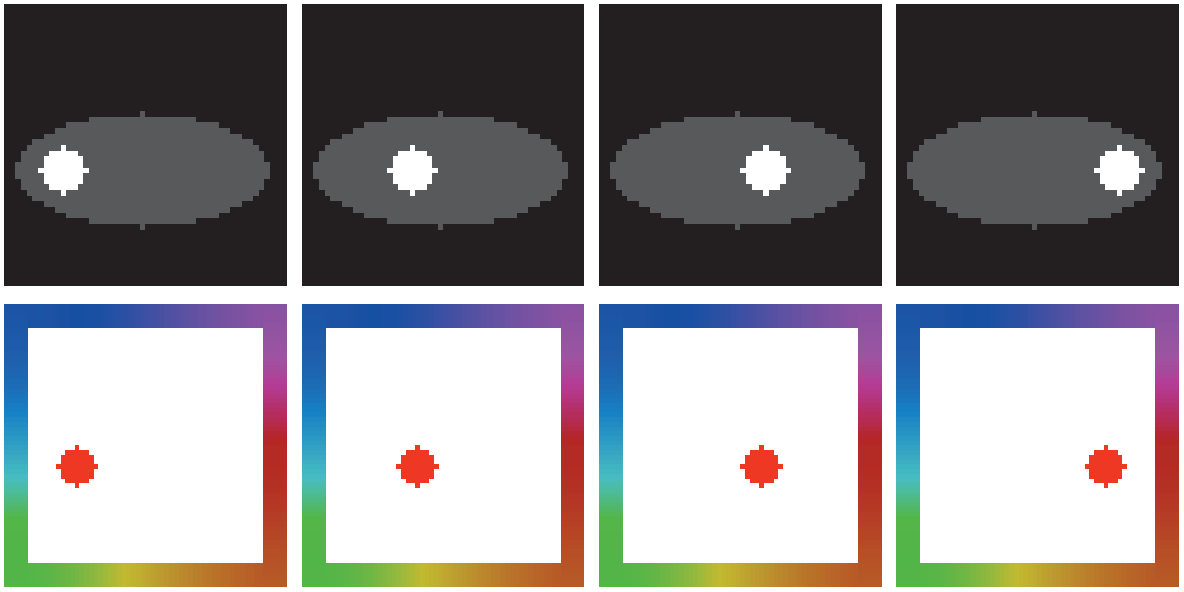

Figure 6.2: Pinball data set. Top: ground truth reconstruction, bottom: ground truth flow field. From left to right: time steps 1, 10, 20, 30 (flow field: 29 instead of 30).

The resulting choices for θ can be found in Table 6.1. We start the numerical approach for this setting by using 30 time steps. In the final part of this section we even reduce the number of time steps to 15 respectively eight.

6.4 Numerical Results

We use two different experimental settings throughout this chapter. The first one is a synthetic data set, which we denote by *Pinball*. We use it in order to be able to measure the capability of the mathematical model by measuring the errors compared to the ground truth data. The second data set, which we refer to as *Rolling Stones*, is a real data set measured in the X-ray lab of the University of Helsinki. With this setting, we test the performance of the model while not having the possibility to fall back on a priori information gained by ground truth data. The framework for the implementation we used throughout this section is based on the code published in Dirks (2015).

6.4.1 Pinball Data Set

The synthetic Pinball data set consists of a two dimensional image of 42×42 pixels. The image is recorded at 30 different time steps. Throughout this period of time, a uniform and rigid ball is slowly moving from the left side of the image frame to the right side. During the whole time, the ball resides in a stationary ellipse of medium intensity. Figure 6.2 shows the ground truth image as well as the ground truth flow field for a selection of time steps. Producing the sinogram based on these images

by simply applying the Radon transform would result in an inverse crime, since we are later using the inverse Radon transform for the reconstruction. To avoid this, we compose the data in a higher resolution, i.e. 84×84. By applying the Radon transform to these images we obtain a preliminary sinogram. To this sinogram we add some positive Gaussian noise and finally downscale it to a size of 63×30, such that it fits to the size of the data that we want to use for the final experiments. The exact appearance of the sinogram depends on the measured angles that are considered. Since this data set is synthetic, we are able to pick every possible combination of angles and try out the most convenient setting. The precise choices are discussed in Section 6.3.

Due to the known ground truth images of this synthetic data set, it is possible to evaluate the performance of the chosen mathematical model by calculating proper errors. Two of the most common error measures are the l_1 distance and the l_2 distance between the reconstructed images and the ground truth images. Thus, we also use both measures to compare the results. However, a disadvantage of their utilization is that they are not neutral with respect to the chosen norms of the model. Naturally, the error in l_1 is smaller for images reconstructed with an L^1 data fidelity, and the error in l_2 is smaller for images reconstructed with an L^2 data fidelity. Since one of the goals of this chapter is to compare the results of both data fidelities, it is necessary to to use an uncommitted technique. For this purpose, the method of our choice is the Structural Similarity (SSIM) Index (cf Brunet et al. (2012)). The SSIM index evaluates an image by comparing the structures of two images with a perception-based model. It includes illumination as well as different contrasts in an image. For two images u_1 and u_2 the SSIM index is given by

$$\widehat{SSIM}(u_1, u_2) = \frac{(2\mu_1\mu_2 + c_1)(2\sigma_{12} + c_2)}{(\mu_1^2 + \mu_2^2 + c_1)(\sigma_1^2 + \sigma_2^2 + c_2)},$$

where μ_1 and μ_2 are the averages of u_1 and u_2, σ_1 and σ_1 are the variances of u_1 and u_2, σ_{12} is the covariance of u_1 and u_2, and c_1 as well as c_2 are constants to prevent a division by zero. To compare the structures of two entire image sequences, we add up the SSIM index for every time step and calculate the average, i.e.

$$SSIM(u_1, u_2) = \frac{1}{T+1} \sum_{t=0}^{T} \widehat{SSIM}\left(u_1(x,t), u_2(x,t)\right). \tag{6.4}$$

In contrast to the errors in l_1 and in l_2, the SSIM index is large for images that are similar to the reference image.

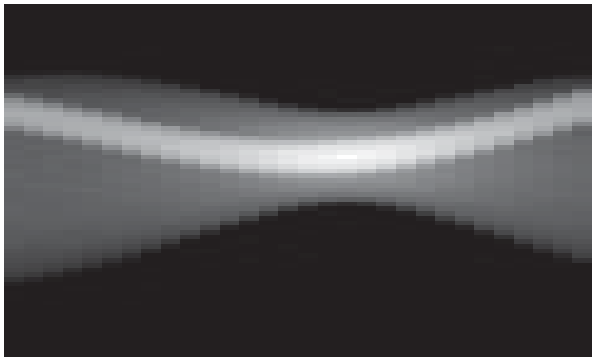

Figure 6.3: Sinogram of Pinball data set for one measured angle per time step.

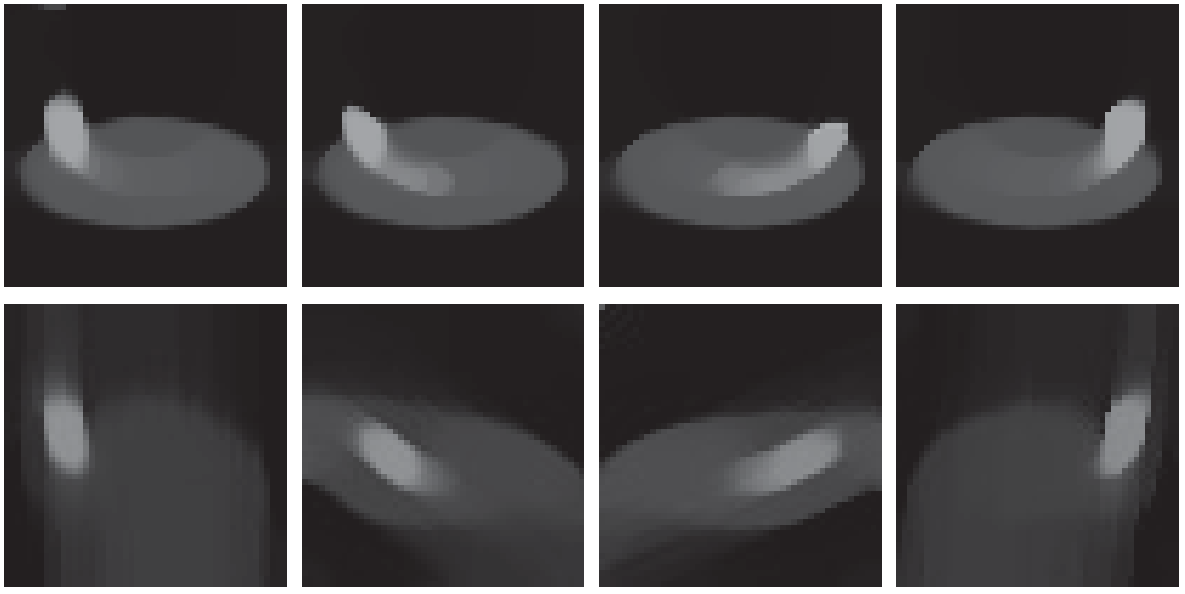

Figure 6.4: Reconstruction result for Pinball data set calculated from one consecutive angle per time step. Top: L^1 data fidelity, bottom: L^2 data fidelity. From left to right: time steps 1, 10, 20, 30.

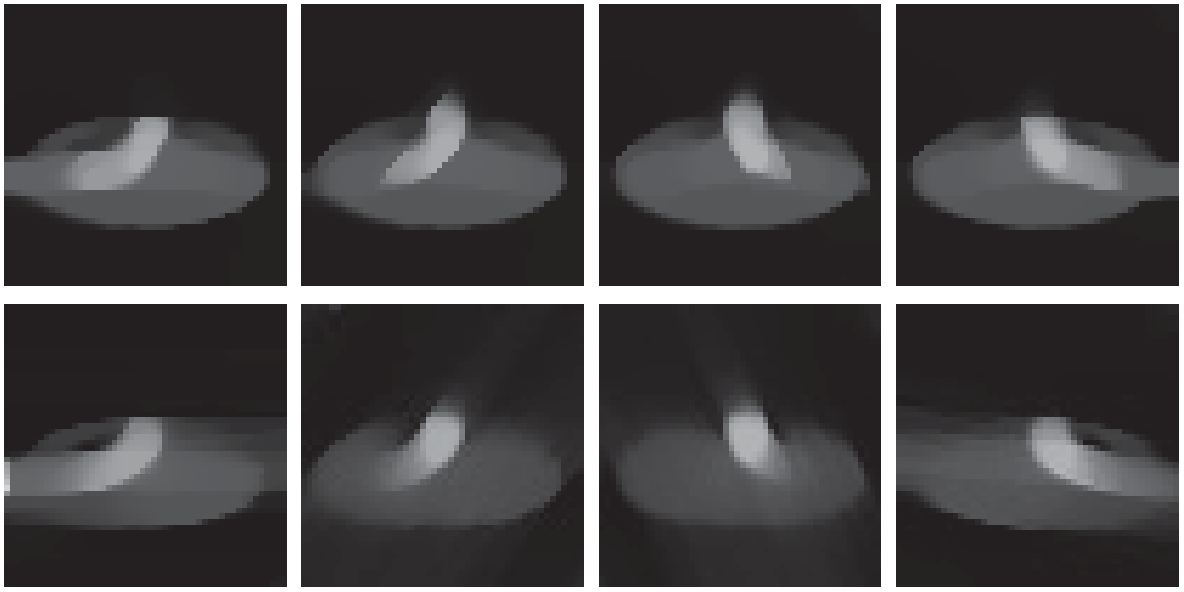

Figure 6.5: Reconstruction result for Pinball data set calculated from one consecutive angle per time step starting at 90°. Top: L^1 data fidelity, bottom: L^2 data fidelity. From left to right: time steps 1, 10, 20, 30.

Figure 6.3 displays a sinogram created with the procedure we described in the beginning of this section. The sinogram contains the measurements from 30 time steps with a single consecutive angle per time step. Hence, this sinogram does not yet consider any of the additional information we introduced in Section 6.3. Figure 6.4 shows a selection of resulting images for the L^1 and for the L^2 data fidelity, which we obtain by applying the indirect model to the sinogram of Figure 6.3. For both data fidelities the ball cannot be reconstructed properly for any of the time steps. Especially in the first and in the last time step the position of the ball is clearly false. It is not located in the ellipse, where it is supposed to be due to the ground truth data. Also the shape of the ball is not correct. The ball follows a wave-like shape, which is a direct consequence of the direction of the actual measured projection. For the L^1 data fidelity the wave is even more visible than for the L^2 data fidelity. In contrast to that, the ellipse in the background is nicely reconstructed with the L^1 data fidelity. The edges are sharp and the ellipse has the correct shape and position for every single time step. The L^2 data fidelity on the other hand reconstructs differing ellipses for the different time steps. In the beginning and in the end, the ellipse is rarely visible. For the middle steps there is an ellipse visible, but it is rather blurry in some areas. Altogether, the results reveal that there is indeed some further a priori information necessary. The calculated reconstructions are the consequence of a local minimum. Calculations with another initialization result in different reconstructions. If the measuring angles are e.g. shifted by 90 degree, the position of the balls seems to follow two smaller waves instead of a single larger one.

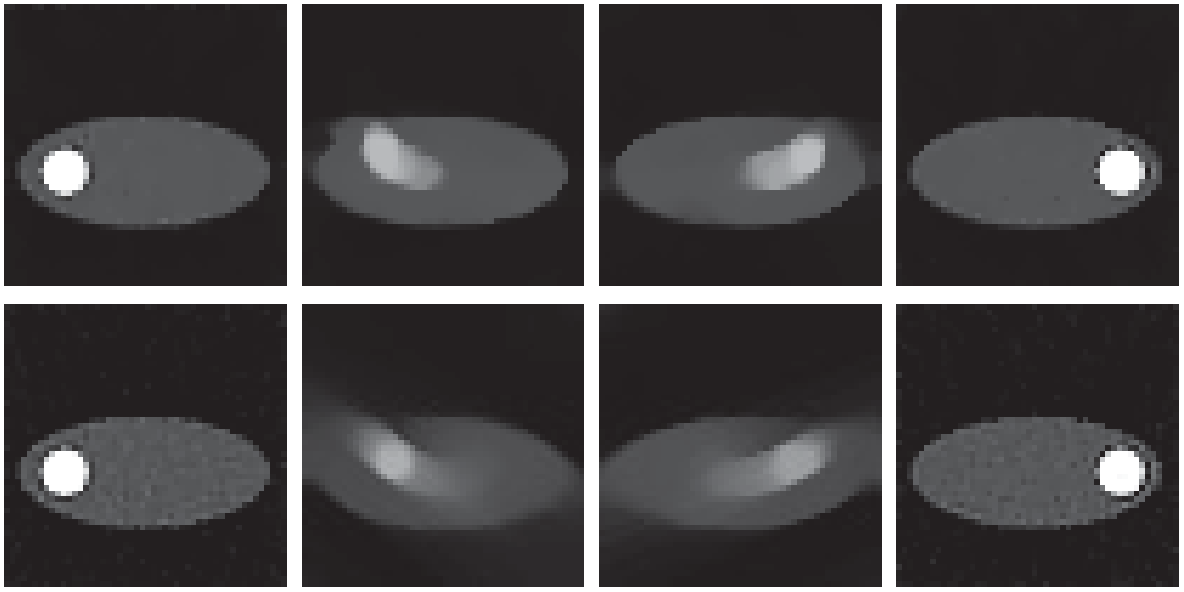

Figure 6.6: Reconstruction result for Pinball data set calculated from one angle per time step with full angle initialization for first and last time step. Top: L^1 data fidelity, bottom: L^2 data fidelity. From left to right: time steps 1, 10, 20, 30.

This effect can be seen in Figure 6.5. The results in this figure are calculated based on the same setting as those in Figure 6.4, except for the starting point of the measurements.

The results for the approach we suggested in Section 6.3.1 are represented in Figure 6.6, again for the L^1 and for the L^2 data fidelity. For the first and the last time step, the ball is well reconstructed for both data fidelities. This is naturally caused by the incorporation of the full 60-angle measurements at these steps. Also the shape of the ellipse is correct. However, the inside of the ellipse has noise-like artifacts for the L^2 data fidelity. For the middle time steps the ball and its position are again not reconstructed correctly. Though, especially the results of the L^1 data fidelity are considerably more accurate than in the approach without further a priori information. The ellipse is again largely properly reconstructed for the L^1 data fidelity, whereas it is still blurry for the L^2 data fidelity. Consequently, this approach, which includes more information for the first and the last time step, does indeed enhance the results, but in particular for the middle steps we need further improvements. The influence of the outer time steps does not reach far enough.

Figure 6.7 depicts the results for the setting, where we measure two orthogonal angles per time step, see Section 6.3.2. In this case, the shape and the position of the ball are already significantly better reconstructed than in the previous results. However, for both data fidelities, but in particular for the L^1 data fidelity, there are two tail-like

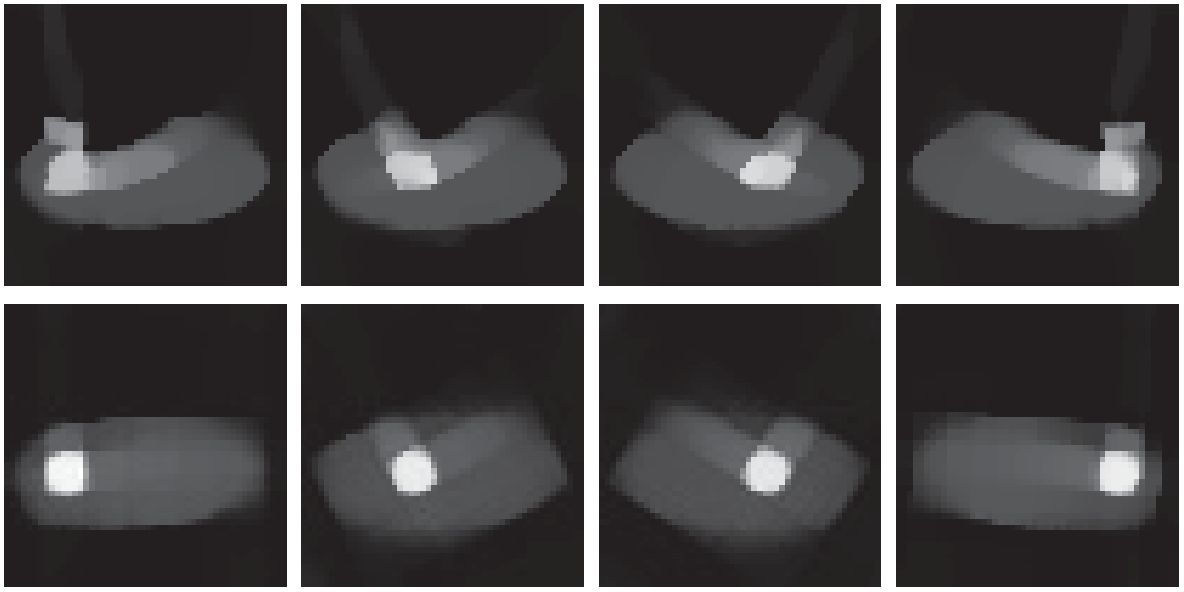

Figure 6.7: Reconstruction result for Pinball data set calculated from two angles per time step with a gap of 90°. Top: L^1 data fidelity, bottom: L^2 data fidelity. From left to right: time steps 1, 10, 20, 29.

artifacts attached to the balls that point in the opposite direction of the actual measured projections. Between those tails, i.e. in the upper part of the results, the ellipses are not properly identified. The results from L^2 data fidelity do only show very few intensity in this area. For the L^1 data fidelity, the model does not reconstruct any intensity at all. The remaining areas of the ellipse have a similar quality as the previous approaches for the L^1 data fidelity. The ellipse reconstructed with the L^2 data fidelity is improved with this setting, but it still suffers from blurry edges.

The results for the approach that includes single randomly distributed angles are represented in Figure 6.8. In this case, the reconstructions as well as the estimated flow fields are shown for each the L^1 and the L^2 data fidelity. The specific angles that are used for this experiment can be found in Table 6.1. In all reconstructions the position of the ball is correctly located. For the L^1 data fidelity the shape of the ball is not completely roundish, but there are no tails visible anymore. Also the edges are rather sharp. For the L^2 data fidelity the general shape of the ball is perfectly reconstructed. However, its edges are still slightly blurry. The ellipse is again well reconstructed for the L^1 data fidelity. The quality of the ellipse of the L^2 data fidelity is clearly better than for all of the previous settings. However, the ellipse of the L^1 data fidelity again looks more accurate due to its sharp boundaries. The pictured flow fields for both data fidelities generally show motion taking place in the correct direction. Nevertheless, there are obvious differences between the flow fields estimated with the L^1 and the L^2 data

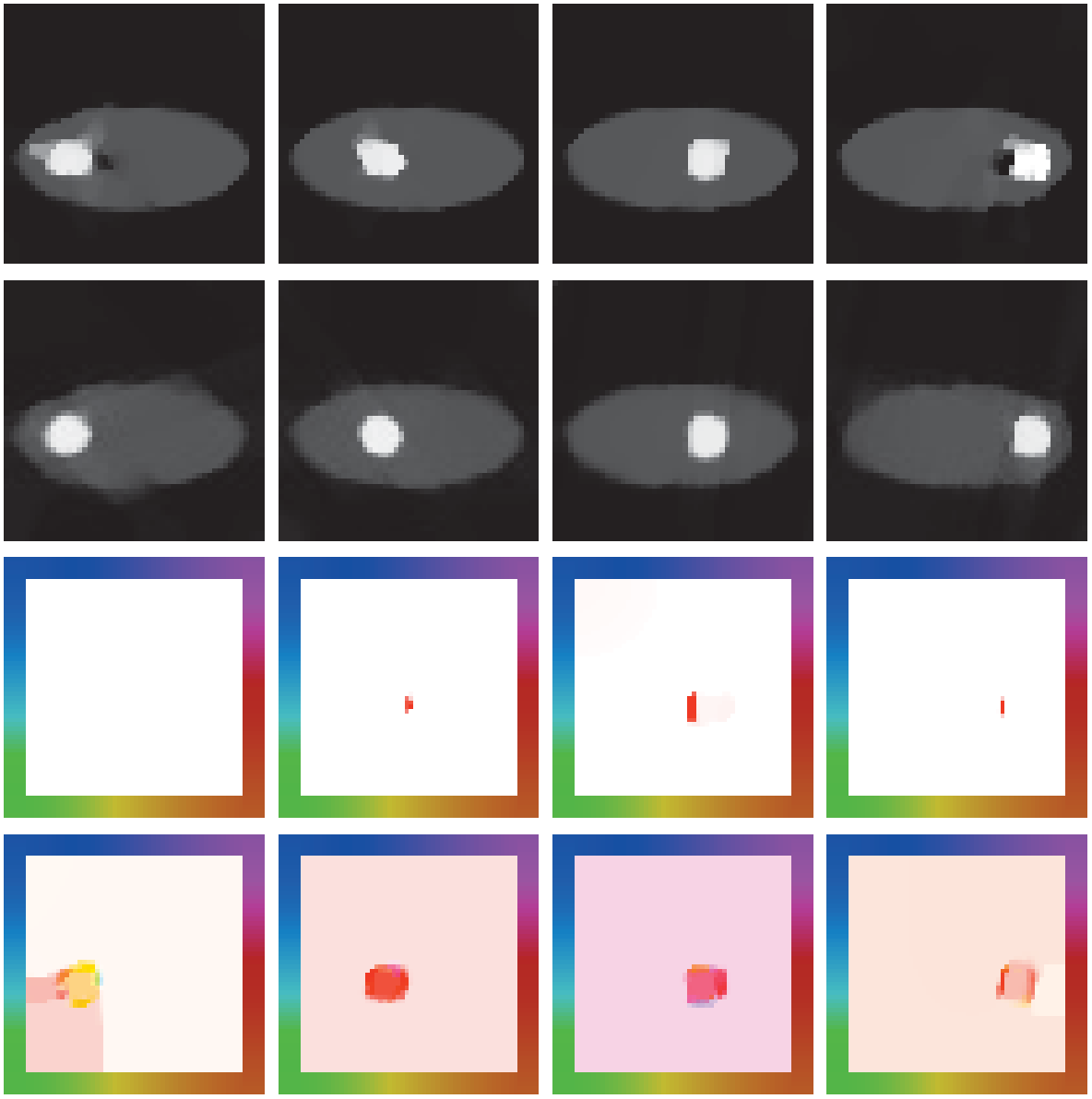

Figure 6.8: Reconstruction and motion estimation result for Pinball data set calculated from one angle per time in a random arrangement. From top to bottom: reconstruction with L^1 data fidelity, reconstruction with L^2 data fidelity, flow field with L^1 data fidelity, flow field with L^2 data fidelity. From left to right: time steps 1, 10, 20, 30 (flow field: 29 instead of 30).

	L^1 data fidelity			L^2 data fidelity		
	ℓ_1 error	ℓ_2 error	SSIM	ℓ_1 error	ℓ_2 error	SSIM
1 angle	0.4744	0.6485	0.7498	0.8897	0.6962	0.4310
1 angle shifted	0.5220	0.7221	0.7163	0.7129	0.6889	0.4855
1 angle + init.	0.3131	0.5177	0.8240	0.4789	0.5042	0.6321
2 angles	0.3828	0.4166	0.7275	0.3329	0.2954	0.7208
1 random angle	0.1978	0.3310	0.8502	0.2223	0.2586	0.8006

Table 6.2: Calculated relative ℓ_1, ℓ_2 errors, and SSIM index for Pinball reconstructions.

fidelity. For the L^2 data fidelity the entire movement of the ball is visible in the flow field. In contrast to that, with the L^1 data fidelity the model only recognizes motion at the outer edges of the ball area. For both data fidelities the flow fields for the first time steps are not correctly estimated. The reason for this is the missing possibility to use a priori information for the calculations from previous time steps, which could be used for initialization.

For all of the previously discussed results, the relative errors in ℓ_1 and in ℓ_2, as well as the SSIM index (see (6.4)) are specified in Table 6.2. Comparing the performances of the L^1 and the L^2 data fidelities for each experimental setting reveals that as expected the error in ℓ_1 is always smaller for the results calculated with the L^1 data fidelity, whereas the ℓ_2 error is smaller for the results of the L^2 data fidelity. However, the SSIM index, which we expect to be more neutral with respect to the chosen norm, indicates that the L^1 norm outperforms the L^2 norm for every single approach. This is a consequence of the distinctly better reconstruction of the ellipses. Concerning the different experimental settings, the approach considering the randomized angles achieves the best results by far. For both data fidelities all error measures indicate that this approach has the best outcome. The order of quality for the remaining settings is however not consistent. For the L^1 data fidelity the second best result is achieved by the approach containing 60-angle measurements for the first and for the last time step. The setting considering only a single measurement at a time is better than the one with two measurements. For the latter the problem is obviously caused by the deficient reconstruction of the ellipse. Regarding the results of the L^2 data fidelity, the setting including two simultaneous measurements produces the second best reconstructions. The approach without any further a priori information results in the largest errors and smallest SSIM index.

6.4.2 Rolling Stones Data Set

The Rolling Stones data set consists of resulting images of the size 42×42, which are each measured at 30 consecutive time steps. Even though we aim at being able to recover continuous movements measured with an extremely limited amount of angles, the measurements were actually recorded from 60 equally distributed angles for every time step. This has the advantage that we are able to use the exact same data set for different arrangements of angles. The following numerical approaches will however not use all of the measured angles. A convenient side effect of this extra measurement is that we are able to use the full 60 angles for reconstructions at every time step to compare them with the later following numerical results. Figure 6.9 shows the reconstruction based on all of the 60 angles for every time step. Even though these are also just approximations of the real objects, we use them similar to a ground truth data set. The reason for this is that results that are based on 60 equally distributed angles are usually quite close to the true object. However, since we will not directly compare these images with the reconstruction results of the indirect model in terms of error measures or the SSIM index, we chose a rather simple method for the 60 angle reconstruction, which includes a smoothed and therefore differentiable TV variant. A detailed description for the used procedure can be found in HÄMÄLÄINEN et al. (2014). The scenery that is represented in the Rolling Stones data set consists of three quadratic stones, which are initially located approximately in the center of the image frame. With every passing time step, the stones move a little bit further apart from each other and closer to different outer boundaries of the frame. Since the stones were moved manually during the measurements, the velocity and direction of movement differs for every stone and every time step. The evaluation of the synthetic data set in the previous section indicates that the L^1 data fidelity works best for the reconstruction of such kind of data. For this reason, we restrict ourselves to this data fidelity for the following results.

Figure 6.10 shows the actual measured sinogram of the setting, which includes a 60-angle initialization for the first and the last time step. The left area in the sinogram up to the first interruption depicts the 60 measurements for the first time step. The middle area illustrates one angle per each of the next 28 steps. The right area shows the 60 measurements for the last time step. The reconstruction results of the indirect model can be seen in Figure 6.11. Similar to the synthetic data set, the first and the last time step is perfectly reconstructed due to the expanded amount of measurements during these steps. However, especially in the seventh time step the stones are not completely separated, which they are actually supposed to be. Even in the 18th step, there is still some blur that connects two of the stones. In general, the position of the stones

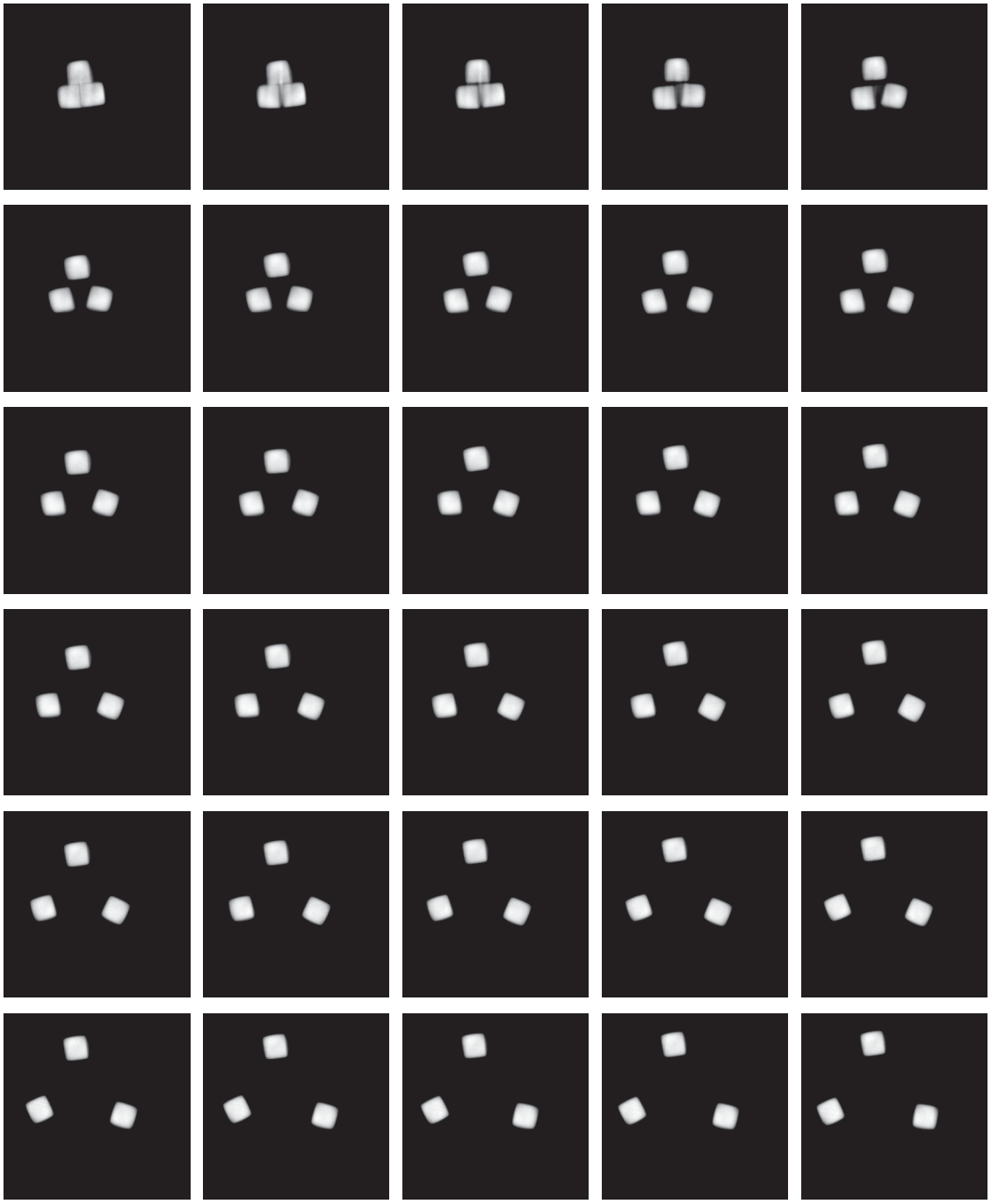

Figure 6.9: Full angle reconstruction of the Rolling Stones data set. From top left to bottom right: time steps 1 to 30.

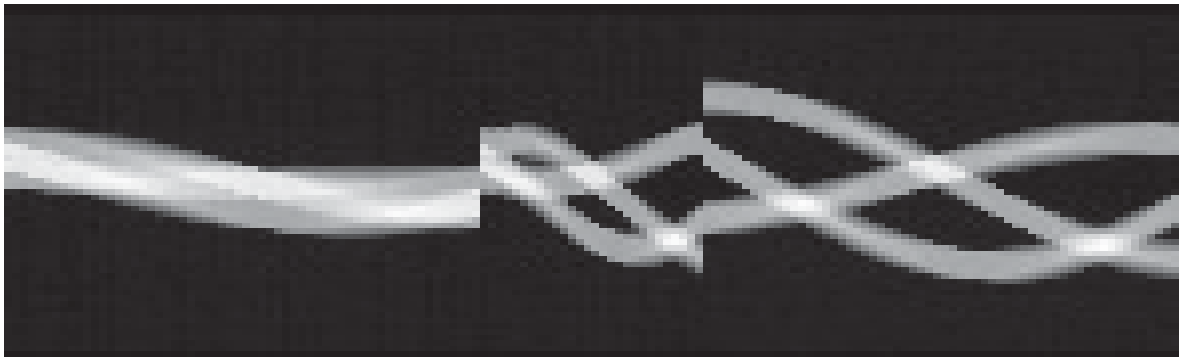

Figure 6.10: Sinogram of Rolling Stones data set for one measured angle per time step and full-angle initialization for first and last time step.

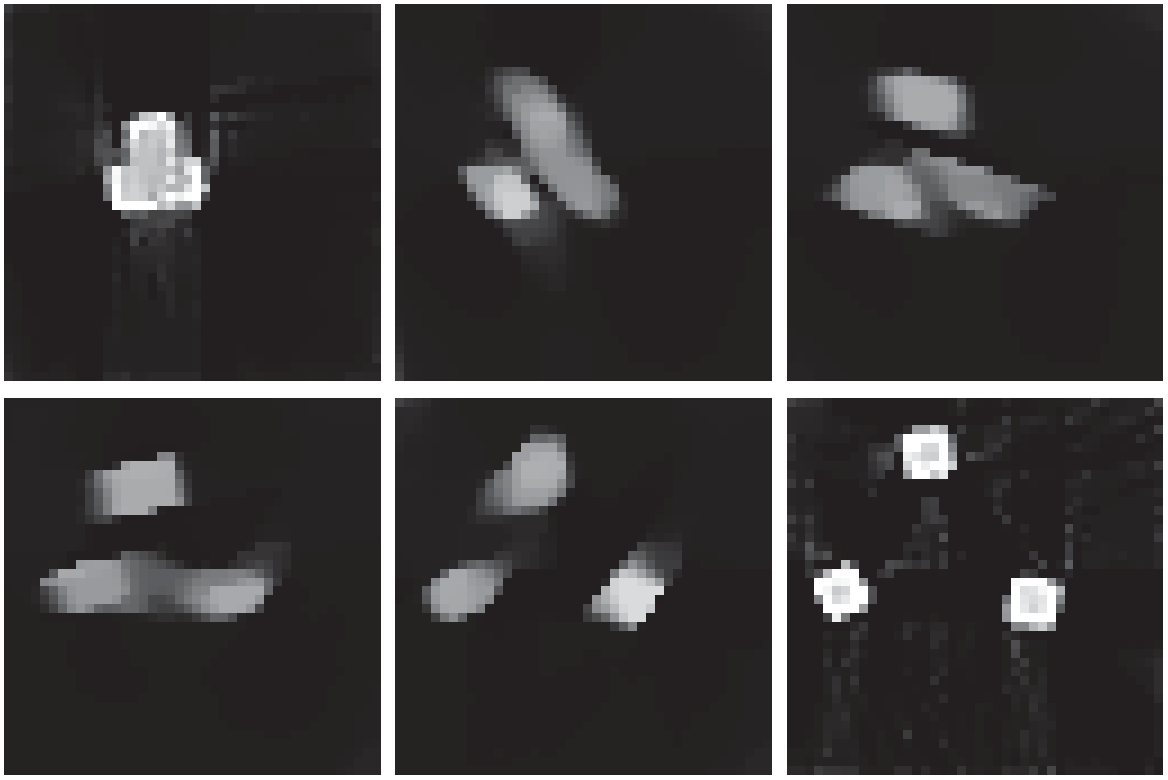

Figure 6.11: Reconstruction result for Rolling Stones data set calculated with L^1 data fidelity from one angles per time step with with full-angle initialization for first and last time step. From top left to bottom right: time steps 1, 7, 13, 18, 25, 30.

is approximately correct, but the shape appears blurry and in some parts without adequate intensity.

The sinogram of the experimental setting regarding two simultaneous angles per time step is illustrated in Figure 6.12. The concurrent time steps are depicted next to each other, such that two sections in the sinogram represent one time step. The results using this sinogram are displayed in Figure 6.13. The first and the last time step are naturally slightly worse than for the previous setting. Nevertheless, both the shape and the position of the stones are very well reconstructed. Also for the middle time steps the images show the stones at the right positions with a distinct separation of the single stones. The only drawback of this approach are sallow artifacts in the seventh time step and some remaining blur in some areas. The latter partly prevents the stones from appearing in the accurate quadratic shape.

For the setting considering 30 randomized angles, the sinogram is demonstrated in Figure 6.14. Again, the specific choice of the angles is given in Table 6.1. The reconstruction results are visualized in Figure 6.15. They show a perfect separation of the stones for all time steps. Also the blur between the single stones is clearly reduced in contrast to the previous approach. The position of the stones is properly reconstructed and in most cases the quadratic shape is well approximated. Even though this result necessitates only 30 measurements in total, which is less than all the other approaches need, it appears to be even more accurate. In this application we aim at reducing the amount of necessary measurements for the dynamic data as far as possible. For this reason, the real data results produced with this randomized approach are already a great success. However, we try to lower the amount of used projections even more to see whether the results are still acceptable. In Figure 6.16 the reconstructions from 15 and from eight measured angles are presented. For the 15 projections, the separation as well as the position of the stones are still appropriately reconstructed. Certainly, there is more blur existent at the boundaries of the stones, which naturally reduces the quality in contrast to the results produced with 30 angles. However, in real applications, where a minimal radiation dose is of particular interest, such results from 15 measurements are indeed reasonable. Regarding the results calculated from only eight measurements, the amount of blur has strongly increased in contrast to the results produced from 15 measurements. The overall shape of the stones is not clear any more. Nevertheless, the stones are still well separated and their position is approximately correct. However, eight seems to be the minimal amount of measured projections in order to be able to produce reasonably useful results.

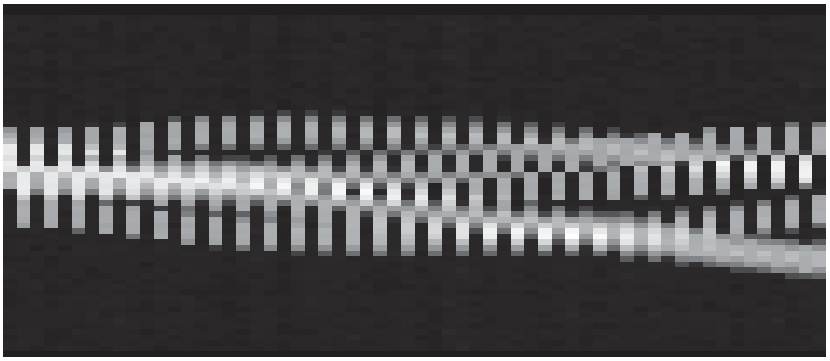

Figure 6.12: Sinogram of Rolling Stones data set for two measured angles per time step.

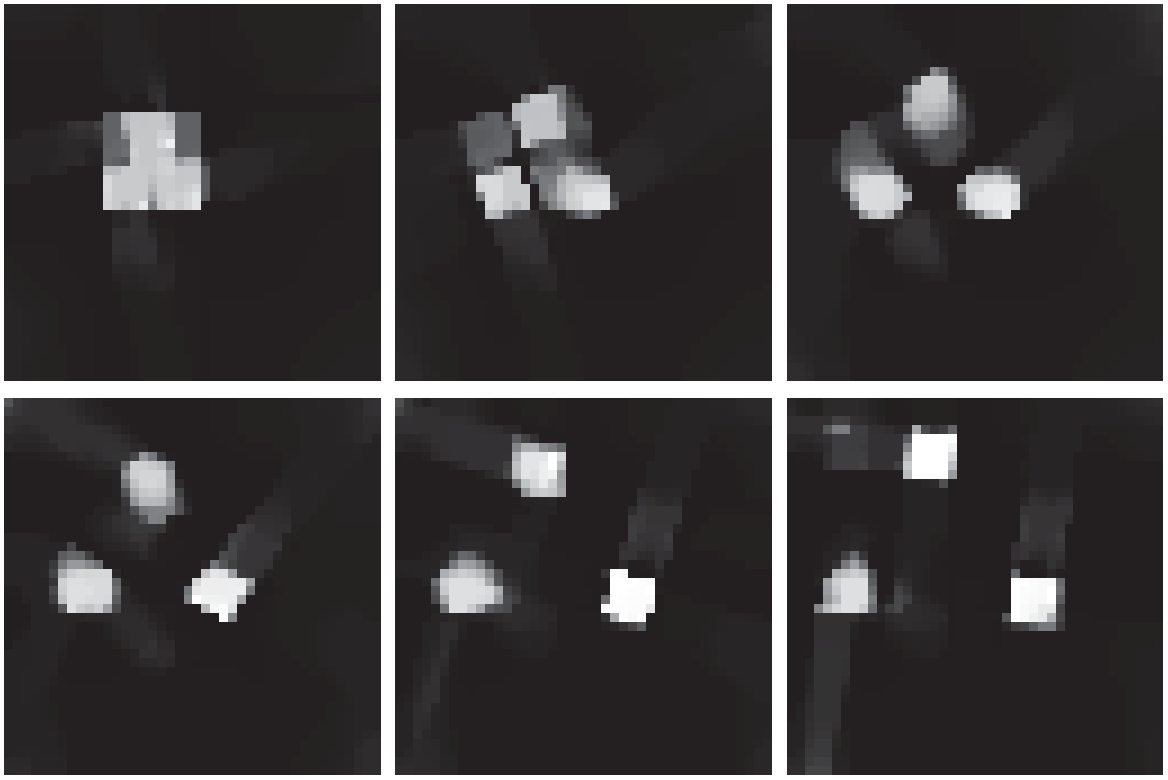

Figure 6.13: Reconstruction result for Rolling Stones data set calculated with L^1 data fidelity from two angles per time step with a gap of 90°. From top left to bottom right: time steps 1, 7, 13, 18, 25, 30.

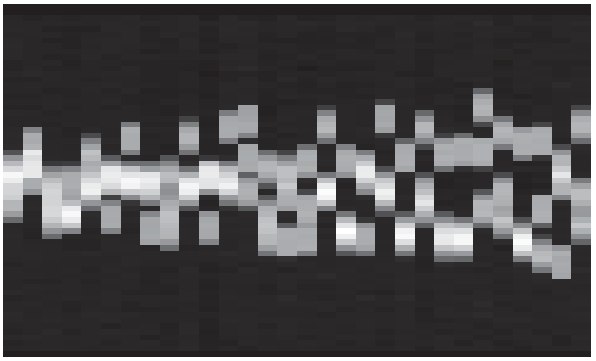

Figure 6.14: Sinogram of Rolling Stones data set for one angle per time in a random arrangement.

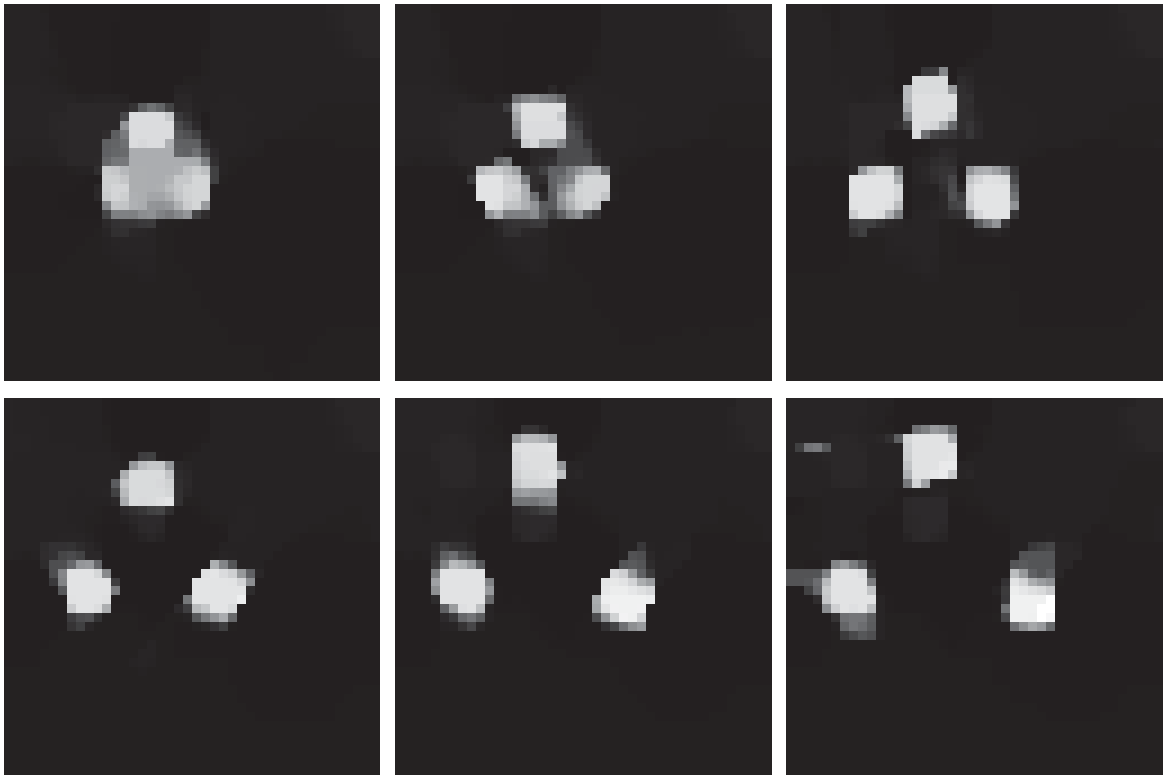

Figure 6.15: Reconstruction result for Rolling Stones data set calculated with L^1 data fidelity from one angle per time in a random arrangement. From top left to bottom right: time steps 1, 7, 13, 18, 25, 30.

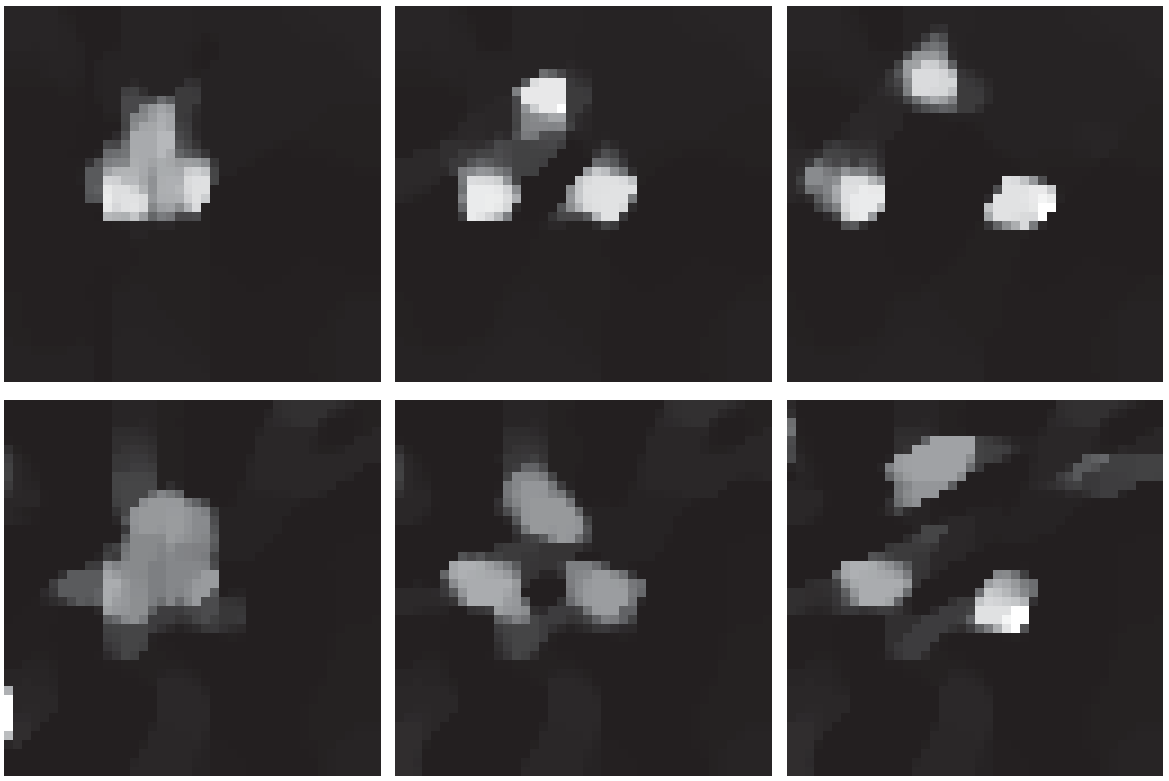

Figure 6.16: Reconstruction result for Rolling Stones data set calculated with L^1 data fidelity from one angle per time in a random arrangement. Top: 15 angles in total, bottom: 8 angles in total. From left to right: time steps 1, 13, 25.

Altogether, the results show that the best performance of the indirect tracking model is achieved by an L^1 data fidelity in a random angle approach with at least 15 measurements. Hence, with this model the reconstruction of dynamic data does not necessarily need more measurements than a reconstruction of static data.

7 Conclusion

Throughout this thesis we studied models for direct and indirect tracking. The latter means that it requires additional reconstruction steps to enable the tracking process. For each type we provided a theoretical as well as a numerical analysis. In order to conclude the thesis, we divide this chapter according to those two types of tracking.

7.1 Direct Tracking

In Chapter 3 we studied various models for motion estimation, all of which are based on an optical flow approach. We introduced data fidelities differing in the optical flow constraint itself, as well as in the applied norm. Concerning the optical flow constraint, we showed a linearized version, which is already well-studied nowadays. However, this variant has the major drawback of only considering very small movements. To overcome this problem, we presented an alternative nonlinear version of the optical flow constraint. We also suggested a way to approximate this constraint and showed the existence of minimizers for this approximation. We studied both the linearized and the nonlinear version in connection with the L^1 and the squared L^2 norm. In addition to the data fidelities, we focused on possibilities to improve the common regularizers. After presenting standard regularizers consisting of only a single term, we proposed an extended regularizer that enables to recover smooth transitions and sharp edges in a single flow field. The general model can be regarded as a further generalization of the TGV regularizer, which is a special case. We proceeded by introducing Bregman iterations for the previously presented models. The remaining part of the chapter was devoted to the numerical realization.

The evaluation of the synthetic results revealed that the performance of the nonlinear

models is superior to the performance of the linearized models, independent of the applied norm. However, we need to mention that this variant depends on significantly more variables and parameters than the linearized variant. Nevertheless, we did not perform any specific optimization for the single models besides of the regularization parameter, which we optimized for both variants. Concerning the regularizers, the results are more ambiguous. Depending on the chosen data set, the extended regularizer are able to improve the results, but it is also possible that a standard regularizer yields slightly better outcomes. Again, the results strongly depend on further parameters linked to the nonlinear data fidelity, which we did not optimize for each single approach. It is certainly possible that the utilized adjustments are simply more suitable for specific combinations of data sets and regularizers. Hence, it is not easily possible to achieve a fair comparison with the nonlinear model. Nevertheless, we chose two of the nonlinear models for the real data application.

In order to estimate the motion of migrating cells based on in vivo microscopy data, we decided to combine the nonlinear optical flow constraint with two variants of the extended regularizer. The results are convincing and show realistic intracellular behavior. However, the results also exhibit small movements in the background of the cells, where there should actually be no visible motion. This characteristic is typical for optical flow models. Since the background in the data is mainly constant, the optical flow constraint is also fulfilled for false detected motion. An identification of pixels based on their intensity is therefore not uniquely possible. A way out of this problem might be to insert a sparsity constraint for the velocity fields. This would achieve a solution with no movements in constant areas.

7.2 Indirect Tracking

In Chapter 4 we discussed the topic of indirect tracking. Since this issue requires a reconstruction step, we introduced a time-dependent reconstruction operator. A combination of this reconstruction with the optical flow models that we derived in the previous section led to a general model for indirect tracking. This model consists of a fidelity to solve the reconstruction task, a regularizer for the reconstructed image, a regularizer for the velocity field, and the optical flow constraint for the motion estimation. We specified two variants for this model, which still have a general reconstruction operator and differ only in the norm of the data fidelity. For these models we also showed the existence of solutions. We additionally discussed the numerical realization, which roughly consists of an alternating minimization that is based on Dirks (2015).

We applied the specified models to a dynamic X-ray tomography approach. For this sake, we shortly discussed the general Radon transform as well as a time-dependent version of it. We used different experimental settings to test the performance of the models. All settings have in common that they only allow for a very limited amount of projections for every time step, whereas the scanned scenery is continuously moving. The best results are achieved by an approach using a random arrangement of angles for consecutive time steps. With this procedure we achieved valuable results with only 15 measurements overall. Hence, with our model the necessary amount of measurements for dynamic data is comparably small as for static data.

Naturally, there are some aspects of the indirect tracking model that remain to be discussed in further studies. The objects we used for scanning had a rather simple shape without any fine structures. Also the movements were relatively slow and straight. The next step in this approach is thus to use more complex data sets and investigate whether fine structures can also be discovered with the indirect tracking model.

The last aspect we suggest for broader analysis is the distribution of the chosen angles. Choosing random angles means that it is also possible to use 15 times the same or very close angles. In this case, it would not be possible to achieve valuable results with any model. Consequently, an open question is how to restrict the distribution of angles in order to obtain the best results.

8

Outlook: Modeling of Motion

This chapter gives an outlook on flow modeling and analysis, which can potentially be used as a priori information for motion estimation models. For this purpose, we depart from imaging and instead have a closer look into mathematical modeling of moving particles. This way we are able to gain important information about the behavior of cells, particles or other individuals, which is known to be a meaningful application for the previously studied motion estimation models. Up to this point, motion estimation models are only based on the information, which is visible in the images. They ignore the fact that behavior of particle systems can often be calculated with a certain probability. Using this information can enhance motion estimation by predicting certain directions of moving objects or characteristic structures in a flow field. This outlook can be regarded as a motivation to close the gap between imaging and mathematical modeling.

With this outlook we provide a systematic study of macroscopic models for transport and diffusion under crowded conditions. We focus on systems of convection-diffusion equations for particle densities rather than hydrodynamic equations for velocities and reaction terms. For the most frequently used approaches, we provide a structured overview of the modeling assumptions and derivations from microscopic stochastic models. Subsequently, we rewrite the models in a unified gradient flow formulation. This allows to obtain macroscopic interpretations of the different model paradigms in terms of mobilities and entropies, as well as a first way of comparing different approaches. Due to this unified presentation, this study also serves as a survey of crowded motion models in literature.

The remainder of this chapter is organized as follows: We start with a short intro-

duction into mathematical modeling of motion. In Section 8.2 we derive five different mathematical models to describe the bidirectional movement of two species. One of them is a mean-field model, two are discrete models, and three are macroscopic models. The discrete models differ in the possibility of two particles to exchange their spatial position. The macroscopic models consist of a model that describes only the pressure exerted by the two species, a second model that additionally considers the influence of the empty spaces, and a third model that describes a nearest-neighbor repulsion between the single particles. In the subsequent section, we develop gradient flow formulations for the previously described models and analyze their behavior in relation to entropy dissipation. This serves as a foundation for the subsequent analytic aspects. In Section 8.4 we study the existence of stationary and transient solutions. We continue with an analysis of the asymptotic regime of low densities, where consistency with standard linear transport-diffusion systems is verified.

8.1 Introduction

Mathematical modeling of motion is a prominent and challenging subject, which is motivated by a whole range of important applications. Meaningful examples are the transport of ions through channels (cf. Gillespie et al. (2002)), intracellular transport processes (cf. Bressloff and Newby (2013)), systems of migrating cells (cf. Burger et al. (2007a) and Hillen and Painter (2009)), animal behavior (cf. Dorigo et al. (1996)), and pedestrian motion (cf. Helbing and Molnar (1995)). Obviously, there are several relevant aspects in such complex systems. An issue, which is typically important for modeling of motion is crowding. However, the implications of specific models for size exclusion effects are still not clear in most cases. For this reason, we give a detailed study of different macroscopic size exclusion models and their mathematical properties in this chapter, with the aim of analyzing which kind of effects can be covered.

In this chapter we develop and analyze different kinds of multicellular models. We make use of the well-known models for a single species, and expand them to models for two different species. Most of the presented models could easily be expanded to models for more than two species. All the models developed in this chapter are one-dimensional, i.e. the species can either move to the left or to the right. A special feature of the models in this chapter is the fact that we allow the particles of each of the two species to move both to the left and to the right. Thus, every single particle is able to move bidirectional.

The perspective of observing the movement of the particles is very different for the presented models. Mainly, we have three kinds of models. These are mean-field models,

discrete models, and macroscopic models. Each of them gives us information about different characteristics. Thus, it depends on the specific research goal, whose framework fits the best. The idea of a mean-field model is based on the movement of ions through an electric field. The main characteristic of mean-field models is the fact that the single particles are influenced by an external field, however not by other single particles directly. The field constitutes an average influence, which replaces the influences of the other particles. One of the advantages is that it is not necessary to take care of the behavior of every single particle to make predictions about the density behavior. In contrast to this, discrete models describe single steps of single particles from one lattice site to another. Every particle is treated as an individual and it is possible to make predictions for the movement of every single one. This is an advantage if the number of cells is small enough. On the other hand, computations and simulations are very expensive. A convenient way to describe the behavior of a whole population of particles is the simulation with macroscopic models. Similar to mean-field models, they do not consider the influence of every single cell, but they calculate densities for every species, which influence each other. This is useful, since often a special behavior of small particles is only noticeable on a macroscopic scale and thus can easier be incorporated into this kind of model.

The main kinds of models treated in this section are macroscopic systems of transport-diffusion equations for a vector $\mathbf{c}(x,t) = (c_1(x,t), \dots, c_M(x,t))$ of densities, which are of the form

$$\frac{\partial c_i}{\partial t} = \nabla \cdot \left(\sum_{j=1}^{M} D_{ij}(\mathbf{c}) \nabla c_j + M_{ij}(\mathbf{c}) \nabla V_j \right). \tag{8.1}$$

Here $\mathbf{D} = (D_{ij})$ are diffusion coefficients and $\mathbf{M} = (M_{ij})$ are mobilities. The transport is driven by potential forces with external potentials V_i, which we assume to be stationary for simplicity. Hence, we are able to discuss the effects of crowding and size exclusion models in a generic way independent of other interactions. We observe that the canonical version of the transport-diffusion equations in absence of size exclusion is given by the Fokker-Planck equation

$$\frac{\partial c_i}{\partial t} = \nabla \cdot \left(D_{ii} (\nabla c_i + c_i \nabla V_i) \right), \tag{8.2}$$

with D_{ii} independent of $\mathbf{c}$. This means that $\mathbf{D}$ is a diagonal matrix with constant entries and $\mathbf{M}$ is a diagonal matrix with entries $D_{ii}c_i$. Thus, we are able to distinguish crowding models by nonlinearities in $\mathbf{D}$ and $\mathbf{M}$ and in the system case by off-diagonal terms in D and M.

8.2 Derivation

In this section we introduce different models for bidirectional transport of two species. We start by developing a mean-field model in Section 8.2.1. In Section 8.2.2 we show two discrete models, which differ in the possibility of two neighboring particles to exchange their spaces. Afterwards, we present three different macroscopic models in Section 8.2.3. Two of them are pressure models, one of which considers empty spaces, the other one does not. The third model is a nearest-neighbor repulsion model that respects the influence of both types of particles in the direct environment of a specific particle.

8.2.1 Mean-Field Model

The first model to derive is the mean-field model. It has a physical relationship and describes the movement of charged ions through an electric field. We use Poisson-Nernst-Planck (PNP) equations, which originally contain an equation for the electric field, and a single equation for the density of each ion type. However, for the sake of simplicity, we assume that the influence of the electrostatic field is zero. Hence, we only consider the equations for the densities of the ion types. Here, and also for all of the following models, we assume to have two different types of particles, which we identify by red color referred to as r, and blue color referred to as b. The particles, which are in this case ions, are influenced by diffusion, the electric field in correlation with the charge of the particle, and interactions between the particles. Our model is based on the model in Burger et al. (2007a). For other related mean-field models cf. Arning (2009), Franek (2011), Horng et al. (2012), Lin and Eisenberg (2012).

G^r denotes a matrix that describes the influence of red particles on a red particle; G^{rb} describes the influence of blue particles on a red particle; G^b describes the influence of blue particles on a blue particles; G^{br} describes the influence of red particles on a blue particle. With this notation, the mean-field equations have the following form:

$$\begin{aligned} \frac{\partial r}{\partial t} &= \nabla \cdot \left(D_r \nabla r + r \nabla (r - G^r * r) + r \nabla (b - G^{rb} * b) \right), \\ \frac{\partial b}{\partial t} &= \nabla \cdot \left(D_b \nabla b + b \nabla (b - G^b * b) + b \nabla (r - G^{br} * r) \right). \end{aligned} \tag{8.3}$$

The first term on the right-hand side of the density equations characterizes the diffusion. The second term considers the influence of external forces, i.e. of the electric field. The remaining terms are interaction terms and describe the influence of particles of the same type (third term), and of particles of the other type (fourth term).

8.2.2 Discrete Jump Process

In this subsection we present two discrete models for the motion of the different particle types r and b. In both cases we assume the particle to move along a lattice with equal sites. It is possible for at most one single particle to be on a specific site at the same time. A particle can either move to an empty neighboring site or stay on the same site. In the first model the particles can additionally exchange the spaces with particles on neighboring sites. This feature is excluded for the second model and thus constitutes the only difference between the two of them. In both cases the movement is additionally influenced by forces that act on the particles. V_r is a force that acts on the red particles, and V_b is a force that acts on the blue particles.

Discrete Jumps with Space Exchange

To set up the master equation for the discrete model that includes space exchanges, we need to define the probabilities for the particles to jump to a neighboring lattice site. We denote by $P^{1\pm}$ the probability that a particle of type r jumps to the next site on the right or on the left and that this site is empty before. The probability that the next site on the right or on the left of a red particle is occupied by a blue particle, with which the red particle changes the space, is described by $P^{2\pm}$. $Q^{1\pm}$ and $Q^{2\pm}$ respectively are the appropriate probabilities for the blue particles. For space x, time t and lattice length h these probabilities have the following form:

$$\begin{aligned}
P^{1\pm}(x,t) &= \left(p_r^1 \mp h q_r^1 \nabla V_r \left(x \pm \frac{h}{2}, t\right)\right) \cdot (1 - r_{x\pm h} - b_{x\pm h}),\\
P^{2\pm}(x,t) &= \left(p_r^2 \mp h q_u^2 \nabla V_r \left(x \pm \frac{h}{2}, t\right)\right) \cdot b_{x\pm h},\\
Q^{1\pm}(x,t) &= \left(p_b^1 \mp h q_b^1 \nabla V_b \left(x \pm \frac{h}{2}, t\right)\right) \cdot (1 - r_{u_{x\pm h}} - b_{x\pm h}),\\
Q^{2\pm}(x,t) &= \left(p_b^2 \mp h q_b^2 \nabla V_b \left(x \pm \frac{h}{2}, t\right)\right) \cdot r_{x\pm h}.
\end{aligned} \tag{8.4}$$

The positive constants p_r^1, p_r^2, p_q^1, and p_q^2 characterize the probability to jump to a space in the absence of external forces. Moreover, the constants q_r^1, q_r^2, q_q^1, and q_q^2 specify the influence of the external forces. The rear factor in these equations depicts the probability that the neighboring site is empty or occupied by a suitable particle.

Since we do not distinguish between different particles of the same type, we do not need to determine probabilities for space exchanges within a single type. We are now able to formulate the master equations, which describe the current state of the distinct types

of particles based on the previous state:

$$\begin{aligned}
r(x,t+dt) = {} & r(x,t)\left(1 - P^{1-}(x,t) - P^{1+}(x,t)\right) \\
& + P^{1-}(x+h)r(x+h,t) + P^{1-}(x-h)r(x-h,t) \\
& - r(x,t)\left(P^{2-}(x) + P^{2+}(x)\right) \\
& + P^{2-}(x+h)r(x+h,t) + P^{2-}(x-h)r(x-h,t) \\
& + b(x,t)\left(Q^{2-}(x) + Q^{2+}(x)\right) \\
& - Q^{2-}(x+h)b(x+h,t) - Q^{2-}(x-h)b(x-h,t), \\
b(x,t+dt) = {} & b(x,t)\left(1 - Q^{1-}(x,t) - Q^{1+}(x,t)\right) \\
& + Q^{1-}(x+h)b(x+h,t) + Q^{1-}(x-h)b(x-h,t) \\
& - b(x,t)\left(Q^{2-}(x) + Q^{2+}(x)\right) \\
& + Q^{2-}(x+h)b(x+h,t) + Q^{2-}(x-h)b(x-h,t) \\
& + r(x,t)\left(P^{2-}(x) + P^{2+}(x)\right) \\
& - P^{2-}(x+h)r(x+h,t) - P^{2-}(x-h)r(x-h,t).
\end{aligned}$$

Taylor expansion up to the second order and subsequent insertion of the probabilities (8.4) leads us to:

$$\begin{aligned}
r(x,t+dt) - r(x,t) = {} & 2h^2 q_r^1 \nabla \cdot (r(1-r-b)\nabla V_r) \\
& + h^2 p_r^1 \nabla \cdot ((1-r-b)\nabla r + r(\nabla r + \nabla b)) \\
& + 2h^2 q_r^2 \nabla \cdot (rb\nabla V_r) + h^2 p_r^2 \nabla \cdot (b\nabla r - r\nabla b) \\
& - 2h^2 b^2 \nabla \cdot (rb\nabla V_b) - h^2 p_b^2 \nabla \cdot (r\nabla b - b\nabla r), \\
b(x,t+dt) - b(x,t) = {} & 2h^2 q_b^1 \nabla \cdot (b(1-r-b)\nabla V_b) \\
& + h^2 p_b^1 \nabla \cdot ((1-r-b)\nabla b + b(\nabla r + \nabla b)) \\
& + 2h^2 q_b^2 \nabla \cdot (rb\nabla V_b) + h^2 p_b^2 \nabla \cdot (r\nabla b - b\nabla r) \\
& - 2h^2 r^2 \nabla \cdot (rb\nabla V_r) - h^2 p_r^2 \nabla \cdot (b\nabla r - r\nabla b).
\end{aligned}$$

We use the following scaling to simplify the equations:

$$\frac{p_r^1}{2} = D_r^1, \quad \frac{p_r^2}{2} = D_r^2, \quad \frac{p_b^2}{2} = D_b^2, \quad \frac{2q_r^1}{p_r^1} = \mu_r^1, \quad \frac{2q_r^2}{p_r^2} = \mu_r^2, \quad \frac{2b^2}{p_b^2} = \mu_b^2, \quad \Delta t = 2h^2,$$

which results in

$$\begin{aligned}\frac{\partial r}{\partial t} &= D_r^1\nabla\cdot\big((1-r-b)\nabla r + r(\nabla r+\nabla b) + \mu_r^1 r(1-r-b)\nabla V_r\big)\\ &\quad + D_r^2\nabla\cdot\big((b\nabla r - r\nabla b) + \mu_r^2 rb\nabla V_r\big)\\ &\quad + D_b^2\nabla\cdot\big((-r\nabla b + b\nabla r) + \mu_b^2 rb\nabla V_b\big)\,,\end{aligned} \tag{8.5}$$

$$\begin{aligned}\frac{\partial b}{\partial t} &= D_b^1\nabla\cdot\big((1-r-b)\nabla b + b(\nabla b+\nabla r) + \mu_b^1 b(1-r-b)\nabla V_b\big)\\ &\quad + D_b^2\nabla\cdot\big((r\nabla b - b\nabla r) + \mu_b^2 rb\nabla V_b\big)\\ &\quad + D_r^2\nabla\cdot\big((-b\nabla r + r\nabla b) + \mu_r^2 rb\nabla V_r\big)\,.\end{aligned} \tag{8.6}$$

A detailed derivation of a similar model including space exchange can be found in Frerking (2010).

Discrete Jumps without Space Exchange

If we eliminate the possibility of two neighboring particles to exchange their spaces, the previous model gets reduced to the model discussed in Schlake (2011). The derivation of the model is analog to the previously described model, but the terms that are responsible for the space exchange, i.e. the ones including probabilities with index 2, have to be neglected. Thus, the model has the following form:

$$\begin{aligned}\frac{\partial r}{\partial t} &= D_r^1\nabla\cdot\big((1-r-b)\nabla r + r(\nabla r+\nabla b) + \mu_r^1 r(1-r-b)\nabla V_r\big)\,,\\ \frac{\partial r}{\partial t} &= D_b^1\nabla\cdot\big((1-r-b)\nabla b + b(\nabla b+\nabla r) + \mu_b^1 b(1-r-b)\nabla V_b\big)\,.\end{aligned} \tag{8.7}$$

Further discrete jump models are amongst others derived in Dyson et al. (2012), Landman and Fernando (2011), Penington et al. (2011).

8.2.3 Macroscopic Hydrodynamic Models

In this section we introduce three macroscopic models for bidirectional transport. The first one is a model that calculates the pressure of two different species and considers different associated velocities. The second model respects the influence of the empty spaces in the area not occupied by one of the two types of particles. The third model is a nearest-neighbor model including repulsive effects of particles in the direct environment.

Pressure

To describe the derivation of the particle densities with their appropriate velocities, we divide each particle type according to their direction of movement. Here r_l are the red particles that move to the left, r_r are the red particles that move to the right, b_l are the blue particles that move to the left, b_r are the blue particles that move to the right. We assume that all left-moving particles have the velocity v_l and all right-moving particles have the velocity v_r independent of their particular color. The red particles move with a diffusion constant $\widetilde{D_r}$ and the blue ones with $\widetilde{D_b}$. We allow that the particles are able to change their direction of movement. The notation for the probabilities of these changes is as follows:

$$
\begin{aligned}
P_l^r &: \text{probability for a change from } r_l \text{ to } r_r \\
P_r^r &: \text{probability for a change from } r_r \text{ to } r_l \\
P_l^b &: \text{probability for a change from } b_l \text{ to } b_r \\
P_r^b &: \text{probability for a change from } b_r \text{ to } b_l
\end{aligned}
$$

We suggest the following equations to describe the development of the particle densities:

$$
\begin{aligned}
\frac{\partial r_l}{\partial t} + \nabla \cdot (v_l r_l) &= \widetilde{D_r} \Delta r - P_l^r r_l + P_r^r r_r, \\
\frac{\partial r_r}{\partial t} + \nabla \cdot (v_r r_r) &= \widetilde{D_r} \Delta r - P_r^r r_r + P_l^r r_l, \\
\frac{\partial b_l}{\partial t} + \nabla \cdot (v_l b_l) &= \widetilde{D_b} \Delta b - P_l^b b_l + P_r^b b_r, \\
\frac{\partial b_r}{\partial t} + \nabla \cdot (v_r b_r) &= \widetilde{D_b} \Delta b - P_r^b b_r + P_l^b b_l.
\end{aligned}
$$

The first terms on the right-hand side characterize the diffusion of the particles. The last two terms specify the changes of the direction of movement. To compare this model to the other models presented in this chapter, we need to summarize both equations for the red particles, and both equations for the blue particles by summing them up:

$$
\begin{aligned}
\frac{\partial r}{\partial t} = \frac{\partial r_l}{\partial t} + \frac{\partial r_r}{\partial t} &= D_r \Delta r - \nabla \cdot (v_l r_l + v_r r_r), \\
\frac{\partial b}{\partial t} = \frac{\partial b_l}{\partial t} + \frac{\partial b_r}{\partial t} &= D_b \Delta b - \nabla \cdot (v_l b_l + v_r b_r),
\end{aligned}
$$

with $D_r = 2\widetilde{D_r}$ and $D_b = 2\widetilde{D_b}$. The terms that specify the change of the direction of movement eliminate each other during the last step.

We also consider a model for the velocities of the different particles. These are influenced by the density pressure, by the flux, by the friction, and by the signal gradient:

$$\begin{aligned}\frac{\partial v_l}{\partial t} + v_l \nabla \cdot v_l &= -\frac{k^2}{r_l + b_l}\nabla \cdot (r + b) + \frac{1}{r_l + b_l}\nabla \cdot (s\nabla \cdot v_l) - \lambda v_l + \nabla V_r, \\ \frac{\partial v_r}{\partial t} + v_r \nabla \cdot v_r &= -\frac{k^2}{r_r + b_r}\nabla \cdot (r + b) + \frac{1}{r_r + b_r}\nabla \cdot (s\nabla \cdot v_r) - \lambda v_r + \nabla V_b,\end{aligned} \tag{8.8}$$

where k^2 is the variance of the velocity distribution, s the viscosity, V_r the density of a signal that affects the velocity of r-particles, and V_b the density of a signal that affects the velocity of b-particles. The first terms on the right-hand side are the pressure terms. They cause that the particles are not able to move very fast in case that the pressure is high. The second terms are flux terms. If the surrounding particles that move in the same direction are faster, a single particle will accelerate. If the surrounding particles are slower, it will also slow down. The third terms involve the friction, and the remaining terms include the influence of a signal gradient, which enforces the movement of a distinct type of cell in one direction. The structure of the presented model is a simplified version of the transport model derived in Frerking (2012).

Pressure Including Empty Spaces

In this section we present a model, which includes the influence of the free spaces on the densities of the red and the blue particles. The model is based on Jackson and Byrne (2002) and Jüngel and Stelzer (2012). For further literature cf. Jungel and Stelzer (2013) and Richardson and King (2007). We assume that the empty space behaves like particles with the density e. We will therefore continue referring to them as empty particles. Since all of the space is filled up with red, blue, or empty particles, we obtain $r + b + e = 1$, which is equivalent to

$$e = 1 - r - b.$$

We suggest that the red particles move with a velocity v_r, the blue particles with v_b, and the empty particles with v_e. Thus, we obtain the following equalities:

$$r_t + \nabla(rv_r) = 0, \qquad b_t + \nabla(bv_b) = 0, \qquad e_t + \nabla(ev_e) = 0. \tag{8.9}$$

Since the empty particle only move as a consequence of the movements of the red and blue particles, we achieve

$$ev_e = -rv_r - bv_b.$$

We consider distinct stresses σ_r, σ_b and σ_e for each of the particles types. Additionally, we assume p to be a common pressure acting on all of the particles. We also respect the influence of the different particle types on each other. Let F_{rb} be the force exerted between red and blue particles, F_{re} the force between red and empty particles, and F_{be} the force between blue and empty particles. For all types we suppose that the force, which a certain type exerts on another one, sums up to zero with the negative force exerted by the second one on the first one. With these assumptions, we obtain the following balance equations:

$$\begin{aligned} \nabla(r\sigma_r) + p\nabla r + F_{rb} + F_{re} &= 0, \\ \nabla(b\sigma_r) + p\nabla b - F_{rb} + F_{be} &= 0, \\ \nabla(e\sigma_e) + p\nabla e - F_{re} - F_{be} &= 0. \end{aligned} \tag{8.10}$$

Now we have a closer look on the stresses σ. For the red and the blue particles, the stresses are composed of the common pressure p, and some isotropic pressure P_r or P_b. For the empty particles, the stress only consists of the common pressure, since there is no isotropic pressure for this type. Consequently, we achieve

$$\sigma_r = -(p + P_r), \qquad \sigma_b = -(p + P_b), \qquad \sigma_e = -p.$$

For the isotropic pressures, we suggest that it increases with the density of the particular particle type, and also with the density of the other types. Thus, we use the following dependencies:

$$P_r = s_r r(1 + \theta b), \qquad P_b = s_b b(1 + \theta r),$$

where s_r, s_b and θ are positive constants. Regarding the forces, we assume that they are proportional to the differences of the appropriate velocities in terms of drag force, and to the particle densities. With these assumptions, we formulate

$$F_{rb} = k(v_b - v_r)rb, \qquad F_{re} = k(v_e - v_r)re, \qquad F_{be} = k(v_e - v_b)be,$$

where k is a positive constant. Inserting the force equations into (8.10), we obtain:

$$\begin{aligned} \nabla(rP_r) + \nabla pr &= F_{rb} + F_{re} \\ &= k\left(-r(b+e)v_r + rbv_b + rev_e\right) \\ &= -krv_r. \end{aligned}$$

Solving the above equation with respect to rv_r results in

$$\begin{aligned} rv_r &= -k^{-1}\left(\nabla(rP_r) + \nabla pr\right) \\ &= -k^{-1}\left(\nabla\left(r^2 s_r(1+\theta b)\right) - \nabla\left(s_r r^3(1+\theta b) + s_b r b^2(1+\theta r)\right)\right) \\ &= -k^{-1}\left((1-r)\nabla\left(s_r r^2(1+\theta b)\right) - \nabla r\left(s_b b^2(1+\theta r)\right)\right). \end{aligned}$$

An analog procedure for the blue particles leads us to

$$bv_b = -k^{-1}\left((1-b)\nabla\left(s_b b^2(1+\theta r)\right) - b\nabla\left(s_r r^2(1+\theta b)\right)\right).$$

Finally, we insert the above equations into (8.9) and achieve

$$\begin{aligned} r_t &= k^{-1}\nabla\cdot\left((1-r)\nabla\left(s_r r^2(1+\theta b)\right) - r\nabla\left(s_b b^2(1+\theta r)\right)\right), \\ b_t &= k^{-1}\nabla\cdot\left((1-b)\nabla\left(s_b b^2(1+\theta r)\right) - b\nabla\left(s_r r^2(1+\theta b)\right)\right). \end{aligned} \tag{8.11}$$

Nearest-Neighbor Repulsion

The third macroscopic model is the nearest-neighbor repulsion model. We assume that every particle is influenced by the red and blue particles in the immediate environment. We define a function F that determines the influence of other particles on a distinct particle and a function H that eliminates the influences of particles outside a radius $R(x,t)$ of the regarded particle. This radius is determined by the number of neighbors in the environment. The development of the particle densities appears as follows:

$$\begin{aligned} 0 = \frac{\partial r}{\partial t} &+ \nabla\cdot\left(r(x)\left(\int F_{rr}\left(\frac{x-y}{\epsilon}\right)H\left(\frac{x-y}{R(x)}\right)r(y)\,dy\right.\right. \\ &\left.\left.+\int F_{rb}\left(\frac{x-y}{\epsilon}\right)H\left(\frac{x-y}{R(x)}\right)b(y)\,dy\right)\right), \\ 0 = \frac{\partial b}{\partial t} &+ \nabla\cdot\left(b(x)\left(\int F_{bb}\left(\frac{x-y}{\epsilon}\right)H\left(\frac{x-y}{R(x)}\right)b(y)\,dy\right.\right. \\ &\left.\left.+\int F_{br}\left(\frac{x-y}{\epsilon}\right)H\left(\frac{x-y}{R(x)}\right)r(y)\,dy\right)\right). \end{aligned}$$

The radius R can be identified by the following condition:

$$\int H\left(\frac{x-y}{R(x)}\right)\left(r(x)+b(x)\right)\,dy = C,$$

where C is fixed by the number of particles of the interaction. We perform a change of variables by locally approximating x via

$$y = x + \epsilon\tilde{y},$$

where $\epsilon > 0$ is supposed to be small. This leads us to

$$\begin{aligned}
0 = \frac{\partial r}{\partial t} + \nabla \cdot \Bigg(r(x) \Bigg(\int F_{rr}(\tilde{y}) H\left(\frac{\tilde{y}}{\frac{R(x+\epsilon\tilde{y})}{\epsilon}} \right) r(x+\epsilon\tilde{y})\, d\tilde{y} \\
+ \int F_{rb}(\tilde{y}) H\left(\frac{\tilde{y}}{\frac{R(x+\epsilon\tilde{y})}{\epsilon}} \right) d\tilde{y} \Bigg) b(x+\epsilon\tilde{y}) \Bigg), \\
0 = \frac{\partial b}{\partial t} + \nabla \cdot \Bigg(b(x) \Bigg(\int F_{bb}(\tilde{y}) H\left(\frac{\tilde{y}}{\frac{R(x+\epsilon\tilde{y})}{\epsilon}} \right) b(x+\epsilon\tilde{y})\, d\tilde{y} \\
+ \int F_{br}(\tilde{y}) H\left(\frac{\tilde{y}}{\frac{R(x+\epsilon\tilde{y})}{\epsilon}} \right) r(x+\epsilon\tilde{y})\, d\tilde{y} \Bigg) \Bigg).
\end{aligned}$$

We continue with further scaling:

$$R = \epsilon\widetilde{R}, \qquad z = \frac{\tilde{y}}{\widetilde{R}}, \qquad d\tilde{y} = \widetilde{R} dz,$$

and obtain

$$\begin{aligned}
0 = \frac{\partial r}{\partial t} + \nabla \cdot \Bigg(r(x) \widetilde{R}^d \Bigg(\int F_{rr}\big(\widetilde{R}z\big) H(z)\, r\big(x+\epsilon\widetilde{R}z\big)\, dz \\
+ \int F_{rb}\big(\widetilde{R}z\big) H(z)\, b\big(x+\epsilon\widetilde{R}z\big)\, dz \Bigg) \Bigg), \\
0 = \frac{\partial b}{\partial t} + \nabla \cdot \Bigg(b(x) \widetilde{R}^d \Bigg(\int F_{bb}\big(\widetilde{R}z\big) H(z)\, b(x+\epsilon\widetilde{R}z)\, dz \\
+ \int F_{br}\big(\widetilde{R}z\big) H(z)\, r\big(x+\epsilon\widetilde{R}z\big)\, dz \Bigg) \Bigg),
\end{aligned}$$

where d is the dimension of movement. We also obtain a scaled version of the condition for the radius $\widetilde{R}$:

$$\widetilde{R^d} \int H(z) \left(r(x) + b(x)\right) \left(x + \epsilon\widetilde{R}z\right) dz = C,$$

where

$$\int H(z) \left(r(x) + b(x)\right) \left(x + \epsilon\widetilde{R}z\right) dz \approx r(x) + b(x).$$

Consequently, for the radius $\widetilde{R}$ is holds

$$\tilde{R} \approx (r(x) + b(x))^{-\frac{1}{d}}.$$

With

$$r\big(x + \epsilon\widetilde{R}z\big) = r(x) + \epsilon\widetilde{R}z\nabla r(x) \qquad \text{and} \qquad b\big(x + \epsilon\widetilde{R}z\big) = b(x) + \epsilon\widetilde{R}z\nabla b(x),$$

we are able to transform the equations for r and b in the following way:

$$\begin{aligned}
0 = &\frac{\partial r}{\partial t} + \nabla \cdot \left(r(x)\tilde{R}^d \left(\int H(z)F_{rr}(\tilde{R}z)\,dz \cdot r(x) + \epsilon\tilde{R}(x) \int H(z)F_{rr}(\tilde{R}z)z\,dz \cdot \nabla r \right.\right. \\
&\left.\left. + \int H(z)F_{rb}(\tilde{R}z)\,dz \cdot b(x) + \epsilon\tilde{R}(x) \int H(z)F_{rb}(\tilde{R}z)z\,dz \cdot \nabla b \right) \right), \\
0 = &\frac{\partial b}{\partial t} + \nabla \cdot \left(b(x)\tilde{R}^d \left(\int H(z)F_{bb}(\tilde{R}z)\,dz \cdot b(x) + \epsilon\tilde{R}(x) \int H(z)F_{bb}(\tilde{R}z)z\,dz \cdot \nabla b \right.\right. \\
&\left.\left. + \int H(z)F_{br}\big(\widetilde{R}z\big)\,dz \cdot r(x) + \epsilon\widetilde{R}(x) \int H(z)F_{br}\big(\widetilde{R}z\big)z\,dz \cdot \nabla r \right) \right).
\end{aligned}$$

Since none of the first integrals in each line of the above equations contains a density term, the particular terms are equal to zero. Furthermore, we are able to formulate the remaining integrals as functions dependent on $\widetilde{R}$:

$$\begin{aligned}
u_r\big(\widetilde{R}\big) &:= \int H(z)F_{rr}\big(\widetilde{R}z\big)z\,dz, & v_r\big(\widetilde{R}\big) &:= \int H(z)F_{rb}\big(\widetilde{R}z\big)z\,dz, \\
u_b\big(\widetilde{R}\big) &:= \int H(z)F_{bb}\big(\widetilde{R}z\big)z\,dz, & v_b\big(\widetilde{R}\big) &:= \int H(z)F_{br}\big(\widetilde{R}z\big)z\,dz.
\end{aligned}$$

Altogether, this results in

$$\begin{aligned}
&\frac{\partial r}{\partial t} + \nabla \cdot \left(\epsilon r(x)\widetilde{R}^{d+1} \left(u_r\big(\widetilde{R}\big) \cdot \nabla r + v_r\big(\widetilde{R}\big) \cdot \nabla b \right) \right) = 0, \\
&\frac{\partial b}{\partial t} + \nabla \cdot \left(\epsilon b(x)\widetilde{R}^{d+1} \left(u_b\big(\widetilde{R}\big) \cdot \nabla b + v_b\big(\widetilde{R}\big) \cdot \nabla r \right) \right) = 0.
\end{aligned} \tag{8.12}$$

To model a useful repulsion between the particles, we assume that the forces between particles are stronger in case that they are closer together. Thus, the forces are dependent on the particular distances and have the following structure:

$$\begin{aligned}
F_{rr}(x-y) &= -\frac{k_{rr}}{(m_r+1)\,|x-y|^{m_r+1}}, & F_{rb}(x-y) &= -\frac{k_{rb}}{(n_r+1)\,|x-y|^{n_r+1}}, \\
F_{br}(x-y) &= -\frac{k_{br}}{(n_b+1)\,|x-y|^{n_b+1}}, & F_{bb}(x-y) &= -\frac{k_{bb}}{(m_b+1)\,|x-y|^{n_b+1}},
\end{aligned}$$

for positive constants k_{rr}, k_{rb}, k_{bb}, k_{br}, m_r, m_b, n_r, and n_b. The idea of the model is mainly based on the model in Murray et al. (2009) and Murray et al. (2012), which can also be seen in Agsten (2012).

8.3 Gradient Flow Formulation

For the derivation of the models shown in Section 8.2, we neglected production or reduction effects of particles of the types r or b. Therefore, we have a mass conservation in every model, i.e.

$$\int_\Omega r(x)\,dx = c_r \qquad \text{and} \qquad \int_\Omega b(x)\,dx = c_b, \tag{8.13}$$

where c_r and c_b are constants.

In the remainder of this section we develop gradient flow formulations for the previously introduced models, i.e. we show that they have the following structure:

$$\frac{\partial}{\partial t}\begin{pmatrix} r \\ b \end{pmatrix} = \nabla \cdot \left(M(r,b)\nabla \cdot \begin{pmatrix} E_r \\ E_b \end{pmatrix} \right),$$

with $M(r,b)$ being a quadratic matrix depending on r and b that is different for every single model. E_r and E_b are the derivations of an entropy functional $E(r,b)$ with respect to r and b. $E(r,b)$ can also slightly vary depending on the distinct model.

We use this formulation to check the single models for entropy dissipation properties. For this aim we investigate whether it is $E(r,b)_t \leq 0$. For the derivation of $E(r,b)$ with respect to t it applies

$$\frac{d}{dt}\left(E(r,b)\right) = E_r r_t + E_b b_t.$$

With the help of the Riesz representation theorem and a partial integration, we obtain

$$\begin{aligned} E_r r_t &= -\int_\Omega M_{11}|\nabla E_r|^2 + M_{12}\nabla E_r \nabla E_b\,dx, \\ E_b b_t &= -\int_\Omega M_{22}|\nabla E_b|^2 + M_{21}\nabla E_r \nabla E_b\,dx. \end{aligned}$$

As a consequence it is

$$\begin{aligned}\frac{d}{dt}\left(E(r,b)\right) &= E_r r_t + E_b b_t \\ &= -\int_\Omega M_{11}|\nabla E_r|^2 + M_{22}|\nabla E_b|^2 + (M_{12}+M_{21})\nabla E_r \nabla E_b \, dx.\end{aligned}$$

In case that $M(r,b)$ is positive definite, it is $E_t(r,b) \leq 0$. Therefore, we have an entropy dissipation for every model, which has the above gradient flow formulation and a positive definite matrix $M(r,b)$. This entropy dissipation serves as a starting point for the subsequent analysis of the models with respect to the existence of stationary solution.

8.3.1 Mean-Field Model

In order to specify a gradient flow structure formulation for the mean-field model (8.3), we assume the interaction matrix G to be symmetric, i.e. $G^{rb} = G^{br}$. This is a useful assumption, since in common particle systems the force between two particles is equal to its counter force. We define an entropy functional $E(r,b)$ as follows:

$$\begin{aligned}E(r,b) = &\int_\Omega D_r r \log r + D_b b \log b + \frac{1}{2}(r^2 + b^2 + 2br)\, dx \\ &- \frac{1}{2}\int_\Omega r(G^r * r) + r(G^{rb} * b)\, dr - \frac{1}{2}\int_\Omega b(G_b * b) + b(G^{rb} * r)\, db.\end{aligned} \tag{8.14}$$

We obtain the following derivatives of the entropy functional with respect to r and b:

$$\begin{aligned}E_r &= D_r \log r + r + b - G^r * r - G^{rb} * b, \\ E_b &= D_b \log b + r + b - G^b * b - G^{rb} * r,\end{aligned}$$

and the corresponding gradients

$$\begin{aligned}\nabla E_r &= \frac{D_r}{r}\nabla r + \nabla r + \nabla b - \nabla(G^r * r) - \nabla(G^{rb} * b), \\ \nabla E_b &= \frac{D_b}{b}\nabla b + \nabla b + \nabla r - \nabla(G^b * b) - \nabla(G^{rb} * r).\end{aligned}$$

For

$$M(r,b) = \begin{pmatrix} r & 0 \\ 0 & b \end{pmatrix} \tag{8.15}$$

we finally achieve the gradient flow structure

$$\frac{\partial}{\partial t}\begin{pmatrix} r \\ b \end{pmatrix} = \nabla \cdot \left(M(r,b) \nabla \cdot \begin{pmatrix} E_r \\ E_b \end{pmatrix} \right).$$

To prove the positive definiteness of $M(r,b)$, it is sufficient to verify that all leading principal minors are positive. The first one is simply r, which is positive by definition. The second one is rb, which is positive due to the positivity of r and b. Thus, the matrix $M(r,b)$ is always positive definite, which means that the entropy is dissipating for the mean-field model.

8.3.2 Discrete Jump Process

Discrete Jumps with Space Exchange

To show the gradient flow structure of the discrete jump model with space exchange (8.5) – (8.6), we define $V_b := -V_r$, as well as $\mu_r^1 = \mu_r^2 = \mu_r$ and $\mu_b^1 = \mu_b^2 = \mu_b$. Again, we start by defining the entropy functional $E(r,b)$:

$$\begin{aligned} E(r,b) = \int_\Omega & (r \log r + b \log b + (1-r-b)\log(1-r-b) \\ & + \mu_r r V_r - \mu_b b V_r)\, dx. \end{aligned} \tag{8.16}$$

The derivatives of the entropy functional with respect to r and b have the following structure:

$$\begin{aligned} E_r &= \log r - \log(1-r-b) + \mu_r V_r, \\ E_b &= \log b - \log(1-r-b) - \mu_b V_r. \end{aligned}$$

We are now able to determine the gradients of these derivatives:

$$\begin{aligned} \nabla E_r &= \frac{\nabla r}{r} + \frac{\nabla(r+b)}{1-r-b} + \mu_r \nabla V_r, \\ \nabla E_b &= \frac{\nabla b}{b} + \frac{\nabla(r+b)}{1-r-b} - \mu_b \nabla V_r. \end{aligned}$$

Transformation of (8.5) leads us to

$$\begin{aligned}\frac{\partial r}{\partial t} =& \nabla \cdot \left(D_r^1 (1-r-b) r \left(\frac{\nabla r}{r} + \frac{\nabla (r+b)}{1-r-b} + \mu_r \nabla V_r \right) \right. \\ &+ (D_r^2 + D_b^2) rb \nabla \cdot \left(\left(\frac{\nabla r}{r} + \frac{\nabla (r+b)}{1-r-b} + \mu_r \nabla V_r \right) \right. \\ &\left. \left. - \left(\frac{\nabla b}{b} + \frac{\nabla (r+b)}{1-r-b} - \mu_b \nabla V_r \right) \right) \right) \\ =& \nabla \cdot \left(D_r^1 (1-r-b) r \nabla E_r + (D_r^2 + D_b^2) rb \nabla \cdot (E_r - E_b) \right) \\ =& \nabla \cdot \left(\left(D_r^1 r (1-r-b) + (D_r^2 + D_b^2) rb \right) \nabla E_r - (D_r^2 + D_b^2) rb \nabla E_b \right).\end{aligned}$$

Analog calculations for (8.6) achieve:

$$\frac{\partial b}{\partial t} = \nabla \cdot \left(\left(D_b^1 b (1-r-b) + (D_r^2 + D_b^2) rb \right) \nabla E_b - (D_r^2 + D_b^2) rb \nabla E_r \right).$$

Thus, we obtain the following gradient flow structure:

$$\frac{\partial}{\partial t} \begin{pmatrix} r \\ b \end{pmatrix} = \nabla \cdot \left(M(r,b) \nabla \cdot \begin{pmatrix} E_r \\ E_b \end{pmatrix} \right),$$

with

$$M(r,b) = \begin{pmatrix} D_r^1 r (1-r-b) + (D_r^2 + D_b^2) rb & -(D_r^2 + D_b^2) rb \\ -(D_r^2 + D_b^2) rb & D_b^1 b (1-r-b) + (D_r^2 + D_b^2) rb \end{pmatrix}. \quad (8.17)$$

To test $M(r,b)$ for positive definiteness, we analyze the leading principal minors. For the first one it applies

$$M_f(r,b) = D_r^1 r (1-r-b) + (D_r^2 + D_b^2) rb,$$

which is positive, since $r+b$ is smaller than one and thus $1-r-b > 0$. The remaining parameters are positive by definition. The second leading principal minor is given by

$$\begin{aligned}M_s(r,b) =& \left(D_r^1 r (1-r-b) + (D_r^2 + D_b^2) rb \right) \left(D_b^1 b (1-r-b) + (D_r^2 + D_b^2) rb \right) \\ &- (D_r^2 + D_b^2)^2 r^2 b^2 \\ =& D_r^1 D_r^1 rb (1-r-b)^2 + (D_r^2 + D_b^2)^2 r^2 b^2 \\ &+ (D_r^1 r + D_b^1 b)(D_r^2 + D_b^2)(1-r-b) rb - (D_r^2 + D_b^2)^2 r^2 b^2 \\ =& D_r^1 D_r^1 rb (1-r-b)^2 + (D_r^1 r + D_b^1 b)(D_r^2 + D_b^2)(1-r-b) rb,\end{aligned}$$

which is greater than zero, since again $(1-r-b)$ as well as all remaining parameters are positive. Consequently, $M(r,b)$ is positive definite and we obtain an entropy dissipation.

Discrete Jumps without Space Exchange

The determination of the gradient flow formulation for the discrete model without space exchange is analog to that for the discrete model with space exchange, which we already showed previously. We choose the same entropy functional as before:

$$\begin{aligned} E(r,b) = \int_\Omega \big(& r\log r + b\log b + (1-r-b)\log(1-r-b) \\ & + \mu_r r V_r - \mu_b b V_r \big) dx. \end{aligned} \tag{8.18}$$

Insertion of the gradients of the derivatives of $E(r,b)$ into (8.7) leads us to:

$$\frac{\partial}{\partial t}\begin{pmatrix} r \\ b \end{pmatrix} = \nabla \cdot \left(M(r,b) \nabla \cdot \begin{pmatrix} E_r \\ E_b \end{pmatrix} \right),$$

with

$$M(r,b) = \begin{pmatrix} D_r^1 r(1-r-b) & 0 \\ 0 & D_b^1 b(1-r-b) \end{pmatrix}. \tag{8.19}$$

Due to the symmetry of the matrix, we check again the positivity of the leading principal minors to verify the positive definiteness of $M(r,b)$. The first leading principal minor is

$$M_f(r,b) = D_r^1 r(1-r-b),$$

which is greater than zero due to the positivity of $(1-r-b)$ and of all remaining parameters. The second leading principal minor is

$$M_s(r,b) = D_r^1 D_b^1 rb(1-r-b)^2,$$

which is also greater than zero. Again, the resulting positive definiteness of $M(r,b)$ gives us an entropy dissipation for this model.

8.3.3 Macroscopic Hydrodynamic Models

Pressure

Before showing the gradient flow structure of the pressure model, we perform a diffusive scaling, i.e.

$$\tilde{t} = \epsilon^2 t, \qquad \tilde{x} = \epsilon x, \qquad \tilde{v} = \frac{v}{\epsilon}.$$

Now it applies

$$\frac{\partial}{\partial \tilde{t}} = \frac{\partial}{\partial t}\left(\frac{\partial t}{\partial \tilde{t}}\right) = \frac{\partial}{\partial t}\frac{1}{\epsilon^2} \qquad \text{and} \qquad \frac{\partial}{\partial \tilde{x}} = \frac{\partial}{\partial x}\left(\frac{\partial x}{\partial \tilde{x}}\right) = \frac{\partial}{\partial x}\frac{1}{\epsilon},$$

and the model for the densities transforms to

$$\begin{aligned}
\epsilon^2 \frac{\partial r(\tilde{x},\tilde{t})}{\partial \tilde{t}} &= \epsilon^2 D_r \Delta r - \epsilon^2 \nabla \cdot (\tilde{v_l} r_l + \tilde{v_r} r_r), \\
\epsilon^2 \frac{\partial b(\tilde{x},\tilde{t})}{\partial \tilde{t}} &= \epsilon^2 D_b \Delta b - \epsilon^2 \nabla \cdot (\tilde{v_l} b_l + \tilde{v_r} b_r).
\end{aligned}$$

After dividing both equations by ϵ^2 and neglecting the tildes, we retrieve the original model:

$$\begin{aligned}
\frac{\partial r}{\partial t} &= D_r \Delta r - \nabla \cdot (v_l r_l + v_r r_r), \\
\frac{\partial b}{\partial t} &= D_b \Delta b - \nabla \cdot (v_l b_l + v_r b_r).
\end{aligned} \tag{8.20}$$

The same scaling as previously used applied to the velocity model (8.8) leads us to

$$\begin{aligned}
\epsilon^3 \frac{\partial \tilde{v_l}}{\partial \tilde{t}} + \epsilon^3 \tilde{v_l} \nabla \cdot \tilde{v_l} &= -\frac{\epsilon}{r_l + b_l} \nabla \cdot (r + b) + \frac{\epsilon^3 s}{r_l + b_l} \Delta \tilde{v_l} - \epsilon v_l, \\
\epsilon^3 \frac{\partial \tilde{v_r}}{\partial \tilde{t}} + \epsilon^3 \tilde{v_r} \nabla \cdot \tilde{v_r} &= -\frac{\epsilon}{r_r + b_r} \nabla \cdot (r + b) + \frac{\epsilon^3 s}{r_r + b_r} \Delta \tilde{v_r} - \epsilon v_r.
\end{aligned}$$

For the above equations, we set $k^2 = \lambda = 1$ and $\nabla V_r = \nabla V_b = 0$. Dividing by ϵ, subsequent approaching of ϵ towards zero and neglecting the tildes yields

$$v_l = -\frac{1}{r_l + b_l} \nabla (r + b) \qquad \text{and} \qquad v_r = -\frac{1}{r_r + b_r} \nabla (r + b),$$

which is equivalent to

$$(r_l + b_l) v_l = -\nabla (r + b) \qquad \text{and} \qquad (r_r + b_r) v_r = -\nabla (r + b).$$

Inserting this into (8.20) results in

$$\begin{aligned}\frac{\partial r}{\partial t} &= D_r \Delta r + \Delta(r+b) = (D_r+1)\Delta r + \Delta b,\\ \frac{\partial b}{\partial t} &= D_b \Delta b + \Delta(r+b) = (D_b+1)\Delta b + \Delta r.\end{aligned} \tag{8.21}$$

We define the following entropy functional:

$$E(r,b) = \int_\Omega r\log r + b\log b\, dx.$$

Subsequently, we are able to determine the gradients of the derivatives with respect to r and b:

$$\nabla E_r = \frac{\nabla r}{r}, \qquad \nabla E_b = \frac{\nabla b}{b}.$$

We insert these gradients into (8.21) and obtain the gradient flow structure

$$\frac{\partial}{\partial t}\begin{pmatrix} r \\ b \end{pmatrix} = \nabla \cdot \left(M(r,b)\nabla \cdot \begin{pmatrix} E_r \\ E_b \end{pmatrix}\right),$$

with

$$M(r,b) = \begin{pmatrix} (D_r+1)r & b \\ r & (D_b+1)b \end{pmatrix}. \tag{8.22}$$

Since the mobility matrix $M(r,b)$ is symmetric, we again check the positivity of the principal leading minors to verify the positive definiteness. The first one is

$$M_f(r,b) = (D_r+1)r,$$

which is greater than zero due to the positivity of the D_r and r. The second one is

$$\begin{aligned}M_s(r,b) &= (D_r+1)(D_b+1)rb - rb\\ &= D_r D_b rb + D_r rb + D_b rb,\end{aligned}$$

which is greater than zero due to the positivity of all parameters. Thus, $M(r,b)$ is positive definite and the entropy is dissipating.

Pressure Including Empty Spaces

Similar to the previous section, for the entropy functional

$$E(r,b) = \int_\Omega r\log r + b\log b\,dx,$$

we obtain the following gradients of the derivatives with respect to r and b:

$$\nabla E_r = \frac{\nabla r}{r}, \qquad \nabla E_b = \frac{\nabla b}{b}.$$

Insertion into (8.11) leads us to

$$\frac{\partial}{\partial t}\begin{pmatrix} r \\ b \end{pmatrix} = \nabla\cdot\left(\widetilde{M}(r,b)\nabla\cdot\begin{pmatrix} E_r \\ E_b \end{pmatrix}\right)$$

with

$$\widetilde{M}(r,b) = \frac{1}{k}\begin{pmatrix} 2s_r r(1-r)(1+\theta b) - s_b\theta r b^2 & s_r\theta r^2 - 2s_b rb(1-\theta r) \\ s_b\theta b^2 - 2s_r rb(1+\theta b) & 2s_b b(1-b)(1+\theta r) - s_r\theta r^2 b \end{pmatrix}.$$

We choose the following scaling:

$$t_s = \frac{t}{\tau}, \qquad x_s = \frac{x}{l}, \qquad \tau = \frac{kl^2}{s_r}, \qquad \beta = \frac{s_b}{s_r},$$

and obtain

$$\frac{\partial}{\partial t_s}\begin{pmatrix} r \\ b \end{pmatrix} = \nabla_{x_s}\cdot\left(M(r,b)\nabla_{x_s}\cdot\begin{pmatrix} E_r \\ E_b \end{pmatrix}\right)$$

with $x_s \in (0,1)$, $t_s > 0$ and

$$M(r,b) = \begin{pmatrix} 2r(1-r)(1+\theta b) - \beta\theta r b^2 & \theta r^2 - 2\beta rb(1-\theta r) \\ \beta\theta b^2 - 2rb(1+\theta b) & 2\beta b(1-b)(1+\theta r) - \theta r^2 b \end{pmatrix}. \tag{8.23}$$

The positive definiteness of this matrix is ambiguous. It strongly depends on the choice of parameters. In case that β is not close to one, i.e. s_r and s_b are very different, the positive definiteness can not be achieved. In case that s_r and s_b are not equal, the external force exerted on one of the particle types is much higher then the force exerted on the other type.

Nearest-Neighbor Repulsion

Once again, we use the entropy functional

$$E(r,b) = \int_\Omega r \log r + b \log b \, dx,$$

and its gradients

$$\nabla E_r = \frac{\nabla r}{r} \qquad \text{and} \qquad \nabla E_b = \frac{\nabla b}{b}.$$

We insert the gradient functions into (8.12) and therefore obtain the gradient flow structure:

$$\frac{\partial}{\partial t}\begin{pmatrix} r \\ b \end{pmatrix} = \nabla \cdot \left(M(r,b) \nabla \cdot \begin{pmatrix} E_r \\ E_b \end{pmatrix} \right),$$

with

$$M(r,b) = \begin{pmatrix} -\epsilon \widetilde{R}^{d+1} r^2(x) u_r\big(\widetilde{R}\big) & -\epsilon \widetilde{R}^{d+1} r(x) b(x) v_r\big(\widetilde{R}\big) \\ -\epsilon \widetilde{R}^{d+1} r(x) b(x) v_b\big(\widetilde{R}\big) & -\epsilon \widetilde{R}^{d+1} b^2(x) u_b\big(\widetilde{R}\big) \end{pmatrix}. \tag{8.24}$$

We assume $F_{rb} = F_{br}$, which means that the force that a particle exerts on another particle is equal to the counterforce of the other particle on the first one. With this assumption the matrix $M(r,b)$ is symmetric and we are able to analyze its positive definiteness by testing the positivity of the principal leading minors. The first one is

$$M_f(r,b) = -\epsilon \widetilde{R}^{d+1} r^2(x) \int H(z) F_{rr}\big(\widetilde{R}z\big) z \, dz,$$

which is greater than zero, since F_{rr} is negative. The second principal leading minor is

$$M_s(r,b) = \epsilon^2 \widetilde{R}^{2(d+1)} r^2(x) b^2(x) \left(\int H(z) F_{rr}\big(\widetilde{R}z\big) z \, dz \int H(z) F_{bb}\big(\widetilde{R}z\big) z \, dz - \left(\int H(z) F_{rb}\big(\widetilde{R}z\big) z \, dz \right)^2 \right).$$

The positivity of this minor depends on the choice of parameters. If we assume F_{rr} and F_{bb} to be greater than F_{rb}, then $M_s(r,b)$, and thus also $M(r,b)$, are positive definite. This choice of parameters means that the forces between two particles of the same type are stronger than the forces between two particles of different types. Only if this condition is fulfilled, the model has a guaranteed entropy dissipation.

8.4 Stationary Solutions

Based on the gradient flow formulations we presented in the previous section, this section is about showing the existence of stationary solutions. We start the investigation by using Lagrange multipliers to determine the critical points. We use (8.13) to obtain the constraints

$$g_r(x) = \int_\Omega r(x,t)\,dx - c_r = 0,$$
$$g_b(x) = \int_\Omega b(x,t)\,dx - c_b = 0,$$

where c_r and c_b are constants. The Lagrangian functional has the form

$$L(r,b) = E(r,b) + \lambda_r \left(\int_\Omega r\,dx - c_r \right) + \lambda_b \left(\int_\Omega b\,dx - c_b \right),$$

with positive constants λ_r and λ_b.

As we already showed in Section 8.3, the mean-field model and the discrete models always have an entropy dissipation. Therefore, critical points of $L(r,b)$ are already stationary solutions for the models. For the macroscopic models we additionally analyze the bordered Hessian to show that the sufficient criterion for a stationary solution is fulfilled.

8.4.1 Mean-Field Model

The entropy functional for the mean-field model is given by

$$\begin{aligned} E(r,b) = \int_\Omega & D_r r \log r + D_b b \log b + \frac{1}{2}(r^2 + b^2 + 2br) + D_r r V_r + D_b b V_b \,dx \\ & - \frac{1}{2}\int_\Omega r(G^r * r) + r(G^{rb} * b)\,dr - \frac{1}{2}\int_\Omega b(G_b * b) + b(G^{rb} * r)\,db. \end{aligned}$$

For the derivatives of the Lagrangian functional it holds

$$\begin{aligned} L_r &= E_r + \lambda_r \\ &= D_r \log r + r + b + D_r V_r - G^r * r - G^{rb} * b + \lambda_r, \\ L_b &= E_b + \lambda_b \\ &= D_b \log b + r + b - G^b * b - G^{rb} * r + \lambda_b. \end{aligned}$$

Critical points for these functionals are existent in case that $L_r = 0$ and $L_b = 0$ hold. To facilitate the analysis of the stationary solutions, we neglect the influence of the electrostatic fields V_r and V_b. Additionally, we assume the logarithm terms to be zero. Hence, the derivatives of the Lagrangian functional reduce to

$$\begin{aligned} L_r &= r + b - G^r * r - G^{rb} * b + \lambda_r, \\ L_b &= r + b - G^b * b - G^{rb} * r + \lambda_b. \end{aligned}$$

Regardless of this reduction, the existence of stationary solutions strongly depends on the appearance of the convolution kernels G^r and G^b. In principle, the above equations could be solved with respect to r and b in case that G^r and G^b are large enough and do not contain zero points. However, in this case they cannot be calculated properly. Consequently, we are not able to state stationary solutions for the mean-field model.

8.4.2 Discrete Jump Process

The discrete models have the entropy functional

$$E(r,b) = \int_\Omega \left(r \log r + b \log b + (1-r-b)\log(1-r-b) + \mu_r r V_r - \mu_b b V_b\right) dx.$$

For the derivatives of the Lagrangian functional with respect to r and b it means that

$$\begin{aligned} L_r &= E_r + \lambda_r \\ &= \log r - \log(1-r-b) + \mu_r V_r + \lambda_r, \\ L_b &= E_b + \lambda_b \\ &= \log b - \log(1-r-b) - \mu_b V_r + \lambda_b. \end{aligned}$$

Since we are searching for critical points, we need to look for points, which satisfy $L_r = 0$, as well as $L_b = 0$. In the absence of the external force V_r, the equations are uniquely solvable for r and b, i.e. we obtain

$$\begin{aligned} r &= \left(1 - \frac{e^{-\lambda_r} + e^{-\lambda_b}}{1 + e^{-\lambda_r} + e^{-\lambda_b}}\right) e^{-\lambda_r}, \\ b &= \left(1 - \frac{e^{-\lambda_r} + e^{-\lambda_b}}{1 + e^{-\lambda_r} + e^{-\lambda_b}}\right) e^{-\lambda_b}. \end{aligned}$$

As previously explained, this critical point is already a stationary solution of $L(r,b)$.

8.4.3 Macroscopic Hydrodynamic Models

All of the macroscopic models have the entropy functional

$$E(r,b) = \int_\Omega r\log r + b\log b\,dx.$$

In order to obtain a critical point, we need to satisfy

$$\begin{aligned} L_r &= E_r + \lambda_r = \log r + \lambda_r = 0,\\ L_b &= E_b + \lambda_b = \log b + \lambda_b = 0. \end{aligned}$$

The unique solutions for these equations are

$$r = e^{-\lambda_r}, \qquad \text{and} \qquad b = e^{-\lambda_b}.$$

To verify a sufficient criterion for stationary solutions we check the sign of the determinant of the Hessian matrix

$$H(r,b) = \begin{pmatrix} 0 & g_r & g_b \\ g_r & L_{rr} & L_{rb} \\ g_b & L_{rb} & L_{bb} \end{pmatrix},$$

where the indices denote the specific derivatives of the particular functional. For the given constraint and the Lagrangian functional we obtain the following determinant of the Hessian matrix $H(r,b)$:

$$\det\left(H(r,b)\right) = \begin{vmatrix} 0 & 1 & 1 \\ 1 & \frac{1}{r} & 0 \\ 1 & 0 & \frac{1}{b} \end{vmatrix} = -\frac{1}{r} - \frac{1}{b}.$$

This determinant is negative for every choice of r and b, since per definition r and b are greater than zero. Thus, it is also negative for the above determined critical point. As a consequence, the critical point is a stationary solution for all of the macroscopic models.

8.5 Small Density Expansion

In this section we perform small density expansions for the different models. Therefore, we approximate the previously calculated mobilities $M(r,b)$ around a small density $\delta = (0,0)$ with the help of Taylor expansions.

8.5.1 Mean-Field Model

We start by analyzing the mobility matrix (8.15) of the mean-field model. The first term of the Taylor expansion at $\delta = (0,0)$ is $M(0,0)$. For the second term, we need to determine the Jacobian matrix of the mobility at $\delta = (0,0)$. The two matrices are given by

$$M(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad J_M(0,0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Hence, the Taylor expansion up to the first order of the mobility matrix reads

$$M(r,b) \approx M(0,0) + J_M(0,0) \begin{pmatrix} r \\ b \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & b \end{pmatrix}.$$

For the Taylor expansion of the entropy functional (8.14), we need to determine $E(0,0)$, the Jacobian matrix at $\delta = (0,0)$, and the Hessian matrix at $\delta = (0,0)$. In this context we assume the diffusion coefficients D_r and D_b to be zero. Hence, logarithm term of the functional can be neglected. With this precondition it holds

$$E(0,0) = c, \qquad J_E(0,0) = \begin{pmatrix} 0 & 0 \end{pmatrix}, \qquad \text{and} \qquad H_E(0,0) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

where c is a constant. Thus, the Taylor expansion up to the second order is given by

$$\begin{aligned} E(r,b) &\approx E(0,0) + J_E(0,0) \begin{pmatrix} r \\ b \end{pmatrix} + \frac{1}{2} \begin{pmatrix} r \\ b \end{pmatrix}^T H_M(0,0) \begin{pmatrix} r \\ b \end{pmatrix} \\ &= c + \frac{1}{2}(r^2 + 2br + b^2). \end{aligned}$$

8.5.2 Discrete Jump Process

Regarding the mobilities (8.17) and (8.19), we see that for the discrete models *with* and *without* space exchange, it holds

$$M(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad J_M(0,0) = \begin{pmatrix} D_r^1 & 0 \\ 0 & D_b^1 \end{pmatrix}.$$

Thus, for both of the discrete models, we obtain the following Taylor expansion at $\delta = (0,0)$:

$$M(r,b) \approx \begin{pmatrix} D_r^1 r & 0 \\ 0 & D_b^1 b \end{pmatrix}.$$

The entropy functionals (8.16) and (8.18), and also their Taylor expansions, are again identical for both of the discrete models. For the remainder of this section, we neglect the terms including $r \log r$ and $b \log b$. Consequently, we achieve

$$E(0,0) = c, \qquad J_E(0,0) = \begin{pmatrix} \mu_r V_r & \mu_b V_b \end{pmatrix}, \qquad \text{and} \qquad H_E(0,0) = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix},$$

where c is a constant. Hence, at $\delta = (0,0)$ we obtain

$$E(r,b) \approx c + \mu_r V_r r + \mu_b V_b b - \frac{1}{2}(r^2 + 2br + b^2).$$

8.5.3 Macroscopic Hydrodynamic Models

For the mobility (8.22) of the pressure model without empty spaces, we have

$$M(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad J_M(0,0) = \begin{pmatrix} D_r + 1 & 0 \\ 0 & D_b + 1 \end{pmatrix}.$$

Altogether, at $\delta = (0,0)$ it holds

$$M(r,b) \approx \begin{pmatrix} (D_r + 1)r & 0 \\ 0 & (D_b + 1)b \end{pmatrix}.$$

For the mobility (8.23) of the pressure model including empty spaces, it applies

$$M(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad J_M(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Therefore, at $\delta = (0,0)$ we obtain

$$M(r,b) \approx \begin{pmatrix} 2r & 0 \\ 0 & 2b \end{pmatrix}.$$

For the mobility (8.24) of the nearest-neighbor repulsion model, we achieve

$$M(0,0) = J_M(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

due to the quadratic density term, and thus at $\delta = (0,0)$ it holds

$$M(r,b) \approx \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The entropy functional of the macroscopic models only consists of logarithm terms. Again, we neglect these terms due to the non-defined Jacobian and Hessian matrices at $\delta = (0,0)$. There remains an integral of zero, which we specify by a constant c, i.e.

$$E(r,b) \approx c.$$

Bibliography

Acar, R. and Vogel, C.R., *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems (1994), 10(6): p. 1217. 37

Agsten, B., *Makroskopische Modelle für Kraftgesetze zwischen Zellen*, Master's thesis, Westfälische Wilhelms-Universität Münster (2012). 178

Ambrosio, L., Fusco, N. and Pallara, D., *Functions of bounded variation and free discontinuity problems*, volume 254, Clarendon Press Oxford (2000). 35

Arning, K., *Mathematical Modelling and Simulation of Ion Channels*, Ph.D. thesis, Radon Institute for Computational and Applied Mathematics (2009). 168

Aubert, G., Deriche, R. and Kornprobst, P., *Computing optical flow via variational techniques*, SIAM Journal on Applied Mathematics (1999), 60(1): pp. 156–182. 19

Aubert, G. and Kornprobst, P., *Mathematical problems in image processing: partial differential equations and the calculus of variations*, volume 147, Springer (2006). 26, 45

Bachmayr, M., *Iterative total variation methods for nonlinear inverse problems*, Master's thesis, Johannes Kepler Universität Linz (2007). 64

Bachmayr, M. and Burger, M., *Iterative total variation schemes for nonlinear inverse problems*, Inverse Problems (2009), 25(10): p. 105004. 64

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M.J. and Szeliski, R., *A database and evaluation methodology for optical flow*, International Journal of Computer Vision (2011), 92(1): pp. 1–31. 86

Barron, J.L., Fleet, D.J. and Beauchemin, S.S., *Performance of optical flow techniques*, International Journal of Computer Vision (1994), 12(1): pp. 43–77. 87

Becker, F., Petra, S. and Schnörr, C., *Optical flow*, Handbook of Mathematical Methods in Imaging (2015), pp. 1945–2004. 45

Bergen, J.R., Anandan, P., Hanna, K.J. and Hingorani, R., *Hierarchical model-based motion estimation*, in *Computer Vision - ECCV'92*, Springer (1992) pp. 237–252. 44

Bertsekas, D.P., *Nonlinear programming*, Athena scientific Belmont (1999). 47

Bian, J., Siewerdsen, J.H., Han, X., Sidky, E.Y., Prince, J.L., Pelizzari, C.A. and Pan, X., *Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam ct*, Physics in medicine and biology (2010), 55(22): p. 6575. 135

Black, M.J., *Robust incremental optical flow*, Ph.D. thesis, Yale university (1992). 46

Blaschke, B., Neubauer, A. and Scherzer, O., *On convergence rates for the iteratively regularized Gauss-Newton method*, IMA Journal of Numerical Analysis (1997), 17(3): pp. 421–436. 47

Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J., *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning (2011), 3(1): pp. 1–122. 40

Bredies, K., Kunisch, K. and Pock, T., *Total generalized variation*, SIAM Journal on Imaging Sciences (2010), 3(3): pp. 492–526. 19, 60

Bredies, K. and Valkonen, T., *Inverse problems with second-order total generalized variation constraints*, Proceedings of SampTA (2011), 201. 60

Bressloff, P.C. and Newby, J.M., *Stochastic models of intracellular transport*, Reviews of Modern Physics (2013), 85(1): p. 135. 166

Brune, C., *4d imaging in tomography and optical nanoscopy*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2010). 20, 40

Brunet, D., Vrscay, E.R. and Wang, Z., *On the mathematical properties of the structural similarity index*, IEEE Transactions on Image Processing (2012), 21(4): pp. 1488–1499. 146

Burger, M., Capasso, V. and Morale, D., *On an aggregation model with long and short range interactions*, Nonlinear Analysis: Real World Applications (2007a), 8(3): pp. 939–958. 166, 168

Burger, M., Dirks, H. and Schönlieb, C.B., *A variational model for joint motion estimation and image reconstruction*, preprint arXiv:1607.03255 (2016a). 100, 110

Burger, M. and Osher, S., *A guide to the TV zoo*, Springer (2013). 35, 61, 107

Burger, M., Papafitsoros, K., Papoutsellis, E. and Schönlieb, C.B., *Infimal convolution regularisation functionals of bv and lp spaces. part i: The finite p case*, Journal of Mathematical Imaging and Vision (2016b), 55(3): pp. 343–369. 59

Burger, M., Resmerita, E. and He, L., *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing (2007b), 81(2-3): pp. 109–135. 61, 97

Chambolle, A., *An algorithm for total variation minimization and applications*, Journal of Mathematical imaging and vision (2004), 20(1-2): pp. 89–97. 85

Chambolle, A. and Pock, T., *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision (2011), 40(1): pp. 120–145. 38, 40

Charbonnier, P., Blanc-Féraud, L., Aubert, G. and Barlaud, M., *Deterministic edge-preserving regularization in computed imaging*, IEEE Transactions on image processing (1997), 6(2): pp. 298–311. 19

Cormack, A.M., *Representation of a function by its line integrals, with some radiological applications*, Journal of applied physics (1963), 34(9): pp. 2722–2727. 136

Cormack, A.M., *Representation of a function by its line integrals, with some radiological applications. ii*, Journal of Applied Physics (1964), 35(10): pp. 2908–2913. 136

Dacorogna, B., *Introduction to the Calculus of Variations*, volume 13, World Scientific (2004). 27, 33

Dirks, H., *Variational Methods for Joint Motion Estimation and Image Reconstruction*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2015). 20, 32, 80, 100, 101, 111, 112, 145, 162

Dirks, H., *https://github.com/hendrikmuenster/motionestimationgui* (2016a). 83

Dirks, H., *Joint large-scale motion estimation and image reconstruction*, preprint arXiv:1610.09908 (2016b). 84, 112

Dorigo, M., Maniezzo, V. and Colorni, A., *Ant system: optimization by a colony of cooperating agents*, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) (1996), 26(1): pp. 29–41. 166

Douglas, J. and Rachford, H.H., *On the numerical solution of heat conduction problems in two and three space variables*, Transactions of the American mathematical Society (1956), 82(2): pp. 421–439. 40

Dyson, L., Maini, P.K. and Baker, R.E., *Macroscopic limits of individual-based models for motile cell populations with volume exclusion*, Physical Review E (2012), 86(3): p. 031903. 171

Ekeland, I. and Témam, R., *Convex analysis and variational inequalities* (1999). 26, 27, 32

Elstrodt, J., *Maß-und Integrationstheorie*, Springer-Verlag (2006). 25

Engl, H.W., Hanke, M. and Neubauer, A., *Regularization of inverse problems*, volume 375, Springer Science & Business Media (1996). 18

Esser, E., Zhang, X. and Chan, T.F., *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences (2010), 3(4): pp. 1015–1046. 38

Evans, L.C., *Partial Differential Equations*, American Mathematical Society (1997). 26, 29, 34

Fleet, D.J. and Jepson, A.D., *Computation of component image velocity from local phase information*, International Journal of Computer Vision (1990), 5(1): pp. 77–104. 87

Forster, O., *Analysis 1*, Springer (2006). 50

Franek, M.M., *Variational Methods using Transport Metrics and Applications*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2011). 168

Frerking, L., *Mathematische Modelle für Chemotaxis gemischter Zellsysteme*, Bachelor's thesis, Westfälische Wilhelms-Universität Münster (2010). 171

Frerking, L., *Mathematical Models for Transport in Axons*, Master's thesis, Westfälische Wilhelms-Universität Münster (2012). 173

Gilland, D.R., Mair, B.A., Bowsher, J.E. and Jaszczak, R.J., *Simultaneous reconstruction and motion estimation for gated cardiac ECT*, IEEE Transactions on nuclear science (2002), 49(5): pp. 2344–2349. 20

Gillespie, D., Nonner, W. and Eisenberg, R.S., *Coupling Poisson–Nernst–Planck and density functional theory to calculate ion flux*, Journal of Physics: Condensed Matter (2002), 14(46): p. 12129. 166

Giusti, E., *Minimal surfaces and functions of bounded variation*, 80, Springer Science & Business Media (1984). 36

Goldstein, T., Li, M., Yuan, X., Esser, E. and Baraniuk, R., *Adaptive primal-dual hybrid gradient methods for saddle-point problems*, preprint arXiv:1305.0546 (2013). 85, 86

Hadamard, J., *Sur les problèmes aux dérivées partielles et leur signification physique*, Princeton university bulletin (1902), 13(49-52): p. 28. 18

Hämäläinen, K., Harhanen, L., Hauptmann, A., Kallonen, A., Niemi, E. and Siltanen, S., *Total variation regularization for large-scale x-ray tomography*, International Journal of Tomography and Simulation (2014), 25(1): pp. 1–25. 153

Hauptmann, A., *Local computerized tomography and total variation regularization*, Master's thesis, Technische Universität München (2012). 138

He, L., Burger, M. and Osher, S.J., *Iterative total variation regularization with non-quadratic fidelity*, Journal of Mathematical Imaging and Vision (2006), 26(1-2): pp. 167–184. 65

Helbing, D. and Molnar, P., *Social force model for pedestrian dynamics*, Physical review E (1995), 51(5): p. 4282. 166

Hillen, T. and Painter, K.J., *A users guide to PDE models for chemotaxis*, Journal of mathematical biology (2009), 58(1-2): pp. 183–217. 166

Horn, B.K. and Schunck, B.G., *Determining optical flow*, in *1981 Technical Symposium East*, International Society for Optics and Photonics (1981) pp. 319–331. 19, 53

Horng, T.L., Lin, T.C., Liu, C. and Eisenberg, B., *PNP equations with steric effects: a model of ion flow through channels*, The Journal of Physical Chemistry B (2012), 116(37): pp. 11422–11441. 168

Huber, P.J., *Robust regression: asymptotics, conjectures and monte carlo*, The Annals of Statistics (1973), pp. 799–821. 57

Huber, P.J. et al., *Robust estimation of a location parameter*, The Annals of Mathematical Statistics (1964), 35(1): pp. 73–101. 57

Jackson, T.L. and Byrne, H.M., *A mechanical model of tumor encapsulation and transcapsular spread*, Mathematical Biosciences (2002), 180(1): pp. 307–328. 173

Jüngel, A. and Stelzer, I.V., *Entropy structure of a cross-diffusion tumor-growth model*, Mathematical Models and Methods in Applied Sciences (2012), 22(07). 173

Jungel, A. and Stelzer, I.V., *Existence analysis of Maxwell–Stefan systems for multicomponent mixtures*, SIAM Journal on Mathematical Analysis (2013), 45(4): pp. 2421–2440. 173

Landman, K.A. and Fernando, A.E., *Myopic random walkers and exclusion processes: Single and multispecies*, Physica A: Statistical Mechanics and its Applications (2011), 390(21): pp. 3742–3753. 171

Ley, K., Laudanna, C., Cybulsky, M.I. and Nourshargh, S., *Getting to the site of inflammation: the leukocyte adhesion cascade updated*, Nature Reviews Immunology (2007), 7(9): pp. 678–689. 120

Lin, T.C. and Eisenberg, B., *A new approach to the Lennard-Jones potential and a new model: PNP-steric equations*, in *2012 International Conference on Nonlinear Analysis: Evoutionary PDE and Kinetic Theory. Taipei, Taiwan: Institute of Mathematics, Academia Sinica* (2012) . 168

Middlebury, *http://vision.middlebury.edu/flow* (2016). 86, 87

Moré, J.J., *The Levenberg-Marquardt algorithm: implementation and theory*, in *Numerical analysis*, pp. 105–116, Springer (1978). 64

Murray, P.J., Edwards, C.M., Tindall, M.J. and Maini, P.K., *From a discrete to a continuum model of cell dynamics in one dimension*, Physical Review E (2009), 80(3): p. 031912. 178

Murray, P.J., Edwards, C.M., Tindall, M.J. and Maini, P.K., *Classifying general nonlinear force laws in cell-based models via the continuum limit*, Physical Review E (2012), 85(2): p. 021921. 178

Natterer, F., *The mathematics of computerized tomography*, volume 32, Siam (1986). 20, 136

Natterer, F. and Wübbeling, F., *Mathematical methods in image reconstruction*, Siam (2001). 139

Niemi, E., Lassas, M. and Siltanen, S., *Dynamic x-ray tomography with multiple sources* (2013), pp. 618–621. 20

Nikolova, M., *Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers*, SIAM Journal on Numerical Analysis (2002), 40(3): pp. 965–994. 45

Orlicz, W., *Beiträge zur Theorie der Orthogonalentwicklungen II*, Studia Mathematica (1929), 1(1): pp. 241–255. 37

Osher, S., Burger, M., Goldfarb, D., Xu, J. and Yin, W., *An iterative regularization method for total variation-based image restoration*, Multiscale Modeling & Simulation (2005), 4(2): pp. 460–489. 20, 61, 62, 63

Otte, M. and Nagel, H.H., *Optical flow estimation: advances and comparisons* (1994), pp. 49–60. 86

Papenberg, N., Bruhn, A., Brox, T., Didas, S. and Weickert, J., *Highly accurate optic flow computation with theoretically justified warping*, International Journal of Computer Vision (2006), 67(2): pp. 141–158. 19

Penington, C.J., Hughes, B.D. and Landman, K.A., *Building macroscale models from microscale probabilistic models: a general probabilistic approach for nonlinear diffusion and multispecies phenomena*, Physical Review E (2011), 84(4): p. 041120. 171

Pettis, B.J., *On integration in vector spaces*, Transactions of the American Mathematical Society (1938), 44(2): pp. 277–304. 37

Radon, J., *1.1 Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten*, Classic papers in modern diagnostic radiology (1917), 5. 20, 136, 138

Ranftl, R., Bredies, K. and Pock, T., *Non-local total generalized variation for optical flow estimation*, in *Computer Vision–ECCV 2014*, pp. 439–454, Springer (2014). 60

Ranftl, R., Gehrig, S., Pock, T. and Bischof, H., *Pushing the limits of stereo using variational stereo estimation*, in *Intelligent Vehicles Symposium (IV), 2012 IEEE*, IEEE (2012) pp. 401–407. 60

Richardson, G. and King, J., *Time-dependent modelling and asymptotic analysis of electrochemical cells*, Journal of Engineering Mathematics (2007), 59(3): pp. 239–275. 173

Rockafellar, R.T., *Convex analysis (princeton mathematical series)*, Princeton University Press (1970), 46: p. 49. 25, 30, 32

Rockafellar, R.T., *Augmented lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of operations research (1976), 1(2): pp. 97–116. 62

Rudin, W., *Functional analysis. 1973*, McGraw-Hill Education (1973). 28

Rudin, W., *Real and complex analysis*, McGraw-Hill Education (1987). 33

Schlake, B.A., *Mathematical Models for Particle Transport: Crowded Motion*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2011). 171

Showalter, R.E., *Monotone operators in Banach space and nonlinear partial differential equations*, volume 49, American Mathematical Soc. (2013). 37, 39

Suhr, S., *Variational Methods for Combined Image and Motion Estimation*, Ph.D. thesis, Westfälische Wilhelms-Universität Münster (2015). 20, 100, 101

Sutton, M., Yan, J., Tiwari, V., Schreier, H. and Orteu, J., *The effect of out-of-plane motion on 2d and 3d digital image correlation measurements*, Optics and Lasers in Engineering (2008), 46(10): pp. 746–757. 42

Vestweber, D., *Adhesion and signaling molecules controlling the transmigration of leukocytes through endothelium*, Immunological reviews (2007), 218(1): pp. 178–196. 121

Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D. and Bischof, H., *Anisotropic Huber-l1 optical flow*, BMVC (2009), 1(2): p. 3. 59

Zach, C., Pock, T. and Bischof, H., *A duality based approach for realtime tv-l 1 optical flow*, in *Pattern Recognition*, pp. 214–223, Springer (2007). 54, 82

Zhang, X., Burger, M. and Osher, S., *A unified primal-dual algorithm framework based on bregman iteration*, Journal of Scientific Computing (2011), 46(1): pp. 20–46. 38