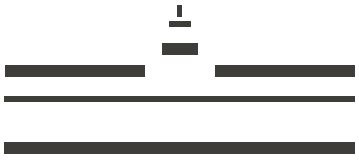
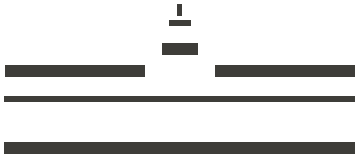


4D Imaging in Tomography and Optical Nanoscopy

Christoph Brune

— 2010 —



Fach Mathematik

4D Imaging in Tomography and Optical Nanoscopy

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

vorgelegt von

Christoph Brune

— 2010 —

Dekan:	Prof. Dr. Christopher Deninger
Erster Gutachter:	Prof. Dr. Martin Burger (Universität Münster)
Zweiter Gutachter:	Prof. Dr. Stanley Osher (Universität von Kalifornien, Los Angeles)
Tag der mündlichen Prüfung:	15.07.2010
Tag der Promotion:	15.07.2010

Acknowledgments

I want to thank

Martin Burger for his mentoring, his belief in me accomplishing this thesis, and his friendship. He gave me a lot of support, by answering my questions, by introducing me to new fields of applied mathematics and by introducing me to many collaborators.

Stanley Osher for reporting my thesis, for meeting with me in Los Angeles last November, and for introducing me to many people of his working group. His ideas on Bregman iteration motivated and helped me a lot.

Eldad Haber for the opportunity of a research stay in Vancouver. I received great hospitality by him and I want to thank him for all he has taught me during that time. Moreover, I want to thank Doug Oldenburg and Uri Ascher for their support.

Helmut Maurer for all he taught me about optimization and optimal control, and for facilitating the first contact to Martin Burger.

my colleagues Martin Benning, Marzena Franek, Michael Möller, Jahn Müller, Lars Ruthotto, Alex Sawatzky, Bärbel Schlake and Frank Wübbeling for advice and fruitful exchange of ideas. Moreover, I thank Martin Benning, Lars Ruthotto, Alexander Wenner and Steffi Brune for carefully reading the manuscript and for giving helpful hints and suggestions. I want to thank all my colleagues in the work group Imaging and in the Institute for Computational and Applied Mathematics for the pleasant working atmosphere.

Marcus Wagner, Karl Kunisch, Yiqiu Dong and Michael Hintermüller for interesting discussions, collaborations and for hosting me in Graz.

Mohammad Dawood, Fabian Gigengack and Xiaoyi Jiang for interesting discussions and collaborations on motion estimation and image registration.

Andreas Schönle and the group of Stefan Hell at the Max-Planck-Institute for biophysical chemistry in Göttingen for collaborations and for providing experimental data.

Paul Lunkenheimer from the university hospital in Münster and the SCANCO Medical AG in Bruettisellen, Switzerland for providing medical imaging data.

the European Institute for Molecular Imaging (EIMI) for the collaboration and for providing data in positron emission tomography.

I acknowledge the Deutsche Telekom Stiftung for their financial support and valuable seminars and contacts. Their financial support made it possible to realize internships and visits at the University of British Columbia in Canada, in Austrian and at the University of California in Los Angeles.

My special thanks go to Christiane Frense-Heck and Reiner Franke for their mentoring and the excellent organization of events.

In particular, I want to thank Torsten Minkwitz from the Deutsche Telekom for his mentoring, for interesting meetings and conversations, and the opportunity to participate in his wide experiences.

I further acknowledge the financial support by the German Federal Ministry of Education and Research through the joint project INVERS: Deconvolution with sparsity constraints in optical nanoscopy and mass spectroscopy under the grant 03BUPAH3.

Finally, I want to thank my family. I am deeply grateful to my wife Steffi, whose love encourages me every day. In my lows she helped me to look on the bright side of life and to get up again. Moreover, I thank my parents, Gisela and Hans, my sister Daniela and my parents in law, Hedwig and Gerhard, for their mental support.

Abstract

This thesis contributes to the field of mathematical image processing and inverse problems. An inverse problem is a task, where the values of some model parameters (in our case images) must be computed from observed data. Such problems arise in a wide variety of applications in sciences and engineering, such as medical imaging, biophysics, geophysics, remote sensing, ocean acoustic tomography, nondestructive testing or astronomy.
Here, we mainly consider reconstruction problems with Poisson noise in tomography and optical nanoscopy. In optical nanoscopy the task is to reconstruct images from blurred and noisy measurements, whereas e.g. in positron emission tomography the task is to visualize biochemical and physiological processes of a patient by measurements from outside the body.
In the literature there are several models and algorithms for 3D static image reconstruction. However, standard methods do not incorporate time-dependent information or dynamics, e.g. heart beat or breathing in medical imaging or cell motion in microscopy. This can lead to deficient accuracy particularly at object boundaries, e.g. at cardiac walls in medical imaging.

This dissertation contains a treatise on models, analysis and efficient algorithms to solve 3D static and 4D time-dependent inverse problems.
In the first part of this thesis the main goal is to present an accurate, robust and fast 3D static reconstruction framework based on total variation for inverse problems with non-standard noise models. We will provide a detailed analysis including existence, uniqueness and convergence proofs.
In the second part our main goal is to study different transport and motion models and to combine them with the ideas of the first part, to build a joint 4D model for simultaneous reconstruction, total variation regularization and optimal transport.

The fundamental concepts are based on non-standard noise models, sparsity regularization techniques, Bregman distances, splitting techniques and motion estimation. In the course of this thesis, topics of various areas in applied mathematics and computer science are brought together, e.g. static and time dependent inverse problems, regularization of ill-posed problems, applied functional analysis, error estimation, convex optimization theory, numerical algorithms, computational science (parallelization, GPU programming), continuum mechanics, computer vision, motion estimation or optimal transport.

The performance of our main concepts is illustrated by experimental data in tomography and optical nanoscopy.

Keywords: 4D image reconstruction, inverse problems, 4D image processing, Poisson noise, total variation, sparsity regularization, Bregman distances, motion estimation, optimal transport, splitting methods, optical nanoscopy, tomography

Contents

Notation and Symbols

Function Spaces and Norms

For an open and bounded subset of $\mathbb{R}^d$ we define the following real-valued function spaces.

$\mathbb{R}^d$	The Euclidean space of dimension d with the Euclidean norm $\lvert\cdot\rvert$.
$\mathbb{R}_+$	Non-negative real numbers.
$BV(\Omega)$	Space of functions of bounded variation with seminorm $\lvert u\rvert_{BV(\Omega)}$, the total variation of u in Ω.
$\mathcal{C}^k(\Omega)$	The space of functions on Ω, which are k-times continuously differentiable.
$L^p(\Omega)$	With $1 \leq p < \infty$: Space of Lebesgue measurable functions u such that $\int_\Omega \lvert u\rvert^p dx < \infty$. The space $L^p(\Omega)$ is a Banach space with corresponding norm $\lVert u\rVert_{L^p(\Omega)} = \left(\int_\Omega \lvert u\rvert^p dx\right)^{\frac{1}{p}}$. In the case $p = 2$ it is a Hilbert space with corresponding inner product $\langle u, v\rangle_{L^2(\Omega)} = \int_\Omega u \cdot v \, dx$.
$\langle \cdot\,, \cdot \rangle_2$	The duality product in $L^2(\Omega)$, $\langle \cdot\,, \cdot \rangle_{L^2(\Omega)}$.
$\lVert\cdot\rVert_2$	The norm in $L^2(\Omega)$, $\lVert\cdot\rVert_{L^2(\Omega)}$.
$L^\infty(\Omega)$	Space of Lebesgue measurable functions u such that there exists a constant C with $\lvert u(x)\rvert \leq C$, a.e. $x \in \Omega$. The space $L^\infty(\Omega)$ is a Banach space with corresponding norm $\lVert u\rVert_{L^\infty(\Omega)} = \sup_{x\in\Omega}\{\, \lvert u(x)\rvert \,\}$.
$L^p_{loc}(\Omega)$	$L^p_{loc}(\Omega) = \{u : \Omega \to \mathbb{R} \mid u \in L^p(D), \text{ for each } D \Subset \Omega\}$.
$W^{k,p}(\Omega)$	With $1 \leq k, p \leq \infty$: Sobolev space of functions $u \in L^p(\Omega)$ such that all weak derivatives up to order k belong to $L^p(\Omega)$. The space $W^{k,p}(\Omega)$ is a Banach space with norm $\lVert u\rVert_{W^{k,p}(\Omega)} = \left(\sum_{i=1}^{k} \int_\Omega \lvert D^i u\rvert^p dx\right)^{\frac{1}{p}}$, where $D^i u$ denotes the i-th distributional derivative of u.
$W^{k,p}_0(\Omega)$	Functions in the Sobolev space $W^{k,p}(\Omega)$ with compact support, i.e. $W^{k,p}_0(\Omega) := \left\{u \in W^{k,p}(\Omega) \mid u\rvert_{\partial\Omega} = 0\right\}$.
$H^k(\Omega)$	$W^{k,2}(\Omega)$. This is a Hilbert space with corresponding inner product $\langle u, v\rangle_{H^k(\Omega)} = \sum_{i=1}^{k} \int_\Omega D^i u \, D^i v \, dx$. For this special Sobolev space we write $\lVert\cdot\rVert_{H^h(\Omega)} := \lVert\cdot\rVert_{W^{k,2}(\Omega)}$ for its corresponding norm.
$\lVert\cdot\rVert_1$	The norm in $H^1(\Omega)$, $\lVert\cdot\rVert_{H^1(\Omega)}$.

$H(div;\Omega)$	The space $H(div;\Omega) := \{\mathbf{v} \in L^2(\Omega)^2 \mid div(\mathbf{v}) \in L^2(\Omega)\}$ becomes an Hilbert space with the inner product $\langle \mathbf{v}, \mathbf{w} \rangle_{div;\Omega} := \langle \mathbf{v}, \mathbf{w} \rangle_\Omega + \langle div\ \mathbf{v}, div\ \mathbf{w} \rangle_\Omega$.
$H(curl;\Omega)$	The space $H(curl;\Omega) := \{\mathbf{v} \in L^2(\Omega)^2 \mid curl(\mathbf{v}) \in L^2(\Omega)\}$ becomes an Hilbert space with the inner product $\langle \mathbf{v}, \mathbf{w} \rangle_{curl;\Omega} \langle \mathbf{v}, \mathbf{w} \rangle_\Omega + \langle curl\ \mathbf{v}, curl\ \mathbf{w} \rangle_\Omega$.

For a Banach space X with a norm $\|\cdot\|_X$ and $\rho : (0,T) \to X$ we denote

$L^p(0,T;X)$	With $1 \leq p < \infty$: Space of functions $\rho \to \rho(t)$ measurable on $(0,T)$ for the measure dt, i.e. the scalar functions $t \to \|\rho\|_X$ are dt-measurable. It is a Banach space with the norm $\|\rho\|_{L^p(0,T;X)} = \left(\int_0^T \|\rho\|_X^p \ dt \right)^{\frac{1}{p}} \leq \infty.$

About Functions and Functionals

For a function $u : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ and a sequence of functions $(u^n)_{n\in\mathbb{N}}$ belonging to a Banach space X we have

$u^n \to u$ in X	The sequence (u^n) converges strongly to u in X.
$u^n \rightharpoonup u$ in X	The sequence (u^n) converges weakly to u in X.
$u^n \rightharpoonup^* u$ in X	The sequence (u^n) converges to u in the weak* topology of X.
$\|u\|_X$	The norm of u in X; for specific norm definitions see previous notation part.
$\mathrm{supp}(u)$	The support of u, $\mathrm{supp}(u) = \Omega \setminus \bigcup_i w_i$, where u is measurable and $(w_i)_{i\in I}$ a family of all open subsets such that $w_i \subseteq \Omega$ and for each $i \in I, u = 0$ a.e. on w_i.
Du	Distributional derivative of u.
∇u	Gradient of u in Ω; we will use $\nabla_\Omega u$ to lay emphasis on Ω.
$\nabla \cdot u$	Divergence of u, i.e. $\nabla \cdot u = \sum_{i=1}^{d} \frac{\partial u}{\partial x_i}$; we will use $\nabla_\Omega \cdot u$ to lay emphasis on Ω.
$\Delta\ u$	Laplacian of u, i.e. $\Delta_\Omega u = \sum_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2}$; we will use $\Delta_\Omega u$ to lay emphasis on Ω.
ρ_t	Temporal derivative $\rho_t = \partial_t \rho$ of a function $\rho : (0,T) \to X$ for $t > 0$.
J^*	The Legendre-Fenchel transform of J, i.e. the convex conjugate $J^* : X^* \to \bar{\mathbb{R}}$ is defined as $J^*(p) = \sup_{u\in X} \{ \langle u, p \rangle - J(u) \}$.

Miscellaneous

Let A, B and R be bounded and open sets in $\mathbb{R}^d$.

$\Omega \subset \mathbb{R}^d$	An open and bounded set with Lipschitz boundary.
$A \hookrightarrow B$	A is continuously embedded into B.
$A \hookrightarrow\hookrightarrow B$	A is compactly embedded into B.
$K \in \mathcal{L}(\mathcal{H})$	K is a bounded linear operator in a Hilbert space $\mathcal{H}$.
K^*	The adjoint operator of K in $\mathcal{H}$, i.e. $\langle Ku, p\rangle = \langle u, K^*p\rangle$, where $\langle \cdot, \cdot \rangle$ denotes the duality product in $\mathcal{H}$.
$\|K\|$	The operator norm of K.
$\|\cdot\|$	Euclidean norm in $\mathbb{R}^d$.
V^*	The topological dual of a topological vector space V.
$\operatorname{sign}(u)$	The sign function, i.e. $\operatorname{sign}(u) = \begin{cases} 1 & u > 0 \\ 0 & u = 0 \\ -1 & u < 0 \, . \end{cases}$
χ_R	Characteristic function of a bounded and open set R, i.e. $\chi_R(u) = \begin{cases} 0 & x \in R \\ +\infty & \text{otherwise} \, . \end{cases}$
$\mathbf{1}_R$	Indicator function of a bounded and open set R, i.e. $\mathbf{1}_R(u) = \begin{cases} 1 & x \in R \\ 0 & \text{otherwise} \, . \end{cases}$

1 Introduction

This thesis deals with models, analysis and algorithms for 3D static and 4D time dependent inverse problems in nanoscopy and tomography. The main idea of this work is to present 3D static reconstruction techniques for inverse problems with non-standard noise models and to simultaneously combine 4D reconstruction algorithms with optimal transport for time dependent inverse problems. The fundamental concepts are based on non-standard noise models, sparsity regularization techniques like total variation (TV), Bregman distances, splitting techniques and motion estimation.

In the course of this thesis, topics of various areas in applied mathematics and computer science are brought together, e.g. mathematical image processing, static and time dependent inverse problems, sparsity regularization of ill-posed problems, applied functional analysis, error estimation, convex optimization theory, numerical algorithms, computational science (parallelization, GPU programming), continuum mechanics, computer vision, motion estimation or optimal transport.

To obtain a good overview of the contributions and the organization of this thesis (Figure 1.2), we will define the main goals of this work in the following. We will start with the motivation of our main problems by considering applications, and then give an outline of the main ideas of this thesis.

1.1 Motivation and Contributions

Inverse Problems and Applications
In the past decades, the field of inverse problems has been a rapidly growing area of applied mathematics. An inverse problem is a mathematical problem where desired model parameters should be reconstructed from observed data. Inverse problems arise in a wide variety of applications in sciences and engineering, such as medical

imaging, biophysics, geophysics, remote sensing, ocean acoustic tomography, nondestructive testing or astronomy.

Inverse problems often are modeled with an operator equation

$$Ku = f \ ,$$

where K denotes a linear (or nonlinear) operator describing the relationship between the observed data f and the desired model parameters u. In this thesis we assume K as a linear and compact operator with a non-closed range, preserving positivity. In imaging u is the desired (density) image we want to reconstruct and f are given measurements possibly defined in another domain. Typical examples for the operator K are Fredholm integral operators of the first kind, i.e.

$$(Ku)(x) = \int_{\Omega} k(x,y)\, u(y)\, dy \ ,$$

where $x \in \Sigma$ and where k denotes a non-negative kernel. Since K cannot be inverted continuously (due to the compactness of the forward operator), most inverse problems are ill-posed following Hadamard.

In this thesis we mainly address the task of reconstructing images from data corrupted by Poisson noise, which is important in various applications, such as fluorescence microscopy, positron emission tomography (PET) or astronomical imaging.

The first application we will consider in this thesis is *optical nanoscopy*. The technology of light microscopy has been considered to be exhausted for a couple of decades, since the resolution is basically limited by Abbe's law for diffraction of light. With conventional microscopy techniques, it is not possible to distinguish features that are located at distances less than about half the wavelength used, i.e. about 200nm for visible light. This diffraction limit is based on the wave nature of light. By developing stimulated emission depletion (STED)- and 4Pi-microscopy [97] nowadays resolutions are achieved way beyond this diffraction barrier [112, 96]. STED-microscopy [191] takes an interesting laser scanning approach, which "practically overcomes" physical limits and allows imaging resolutions in nano-scale. The underlying principle was invented and experimentally realized by Stefan W. Hell and his Max Planck Institute for Biophysical Chemistry in Göttingen.

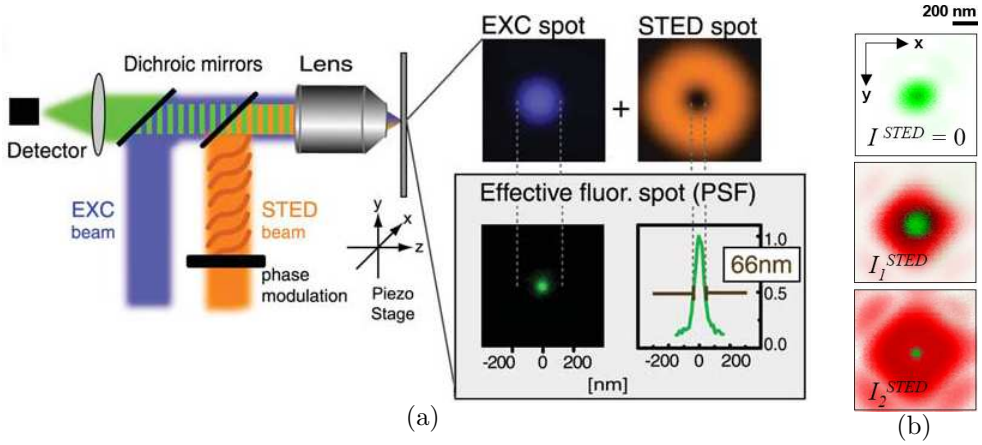


Figure 1.1: Basic principle of a STED microscope: (a) Two laser beams are combined to increase the resolution; (b) Fluorescent spot for different laser intensities of the depletion beam.

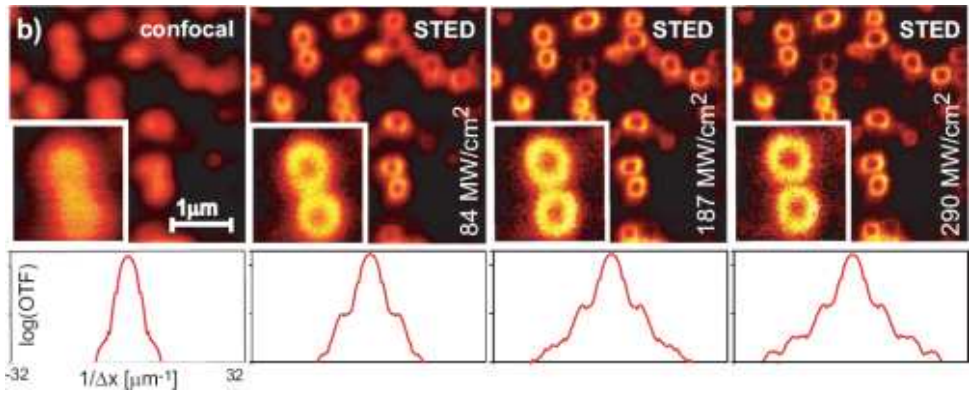


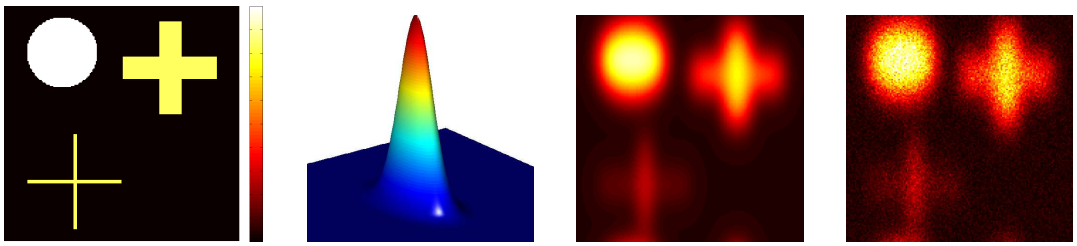
Figure 1.2: Comparison of a confocal light microscope and a STED microscope for different laser intensities; bottom: corresponding PSF structures in the Fourier space.

The basic apparatus consists of two laser beams, a stimulation beam and a depletion beam. Fluorescent dyes are stimulated by the excitation beam ("on-beam") and are directly quenched by an additional interfering laser spot ("off-beam"), the STED spot. Since this depletion beam vanishes at a very small point in the middle, fluorescence of the simulating spot is only detected at this tiny position, see Figure 1.1 (a). For higher laser intensities of the depletion spot the radius of the resulting fluorescence spot reduces, see Figure 1.1 (b) and Figure 1.2. Hence, data with previously unknown resolution can be measured. However, by reaching the diffraction limit of light, measurements f suffer from blurring effects and in addition suffer from Poisson noise due to laser sampling (photon counts).

Consequently, in the case of optical nanoscopy the linear and compact operator K describes a blurring operator, i.e. a convolution operator with a kernel $k \in \mathcal{C}(\Omega \subset \mathbb{R}^d)$

$$(Ku)(x) \;=\; (k * u)(x) \;:=\; \int_\Omega k(x-y)\, u(y) dy \;.$$

The kernel is often referred to as the point spread function (PSF). Our image reconstruction task is to compute deblurred and nearly noise-free images from the measurements of the high resolution fluorescence microscope, see Figure 1.3 for a synthetic illustration and Figure 1.5(a) for a real data set with the protein syntaxin.



desired exact data u | kernel k (psf) | blurred data | given blurred & noisy data f

Figure 1.3: Illustration of the inverse problem setting with synthetic data.

The second application we will consider in this thesis is *tomography in medical imaging*. Positron emission tomography (PET) is a biomedical imaging technique, which enables to visualize biochemical and physiological processes, such as glucose metabolism, blood flow, or receptor concentrations, see e.g. [189], [178], [10]. This

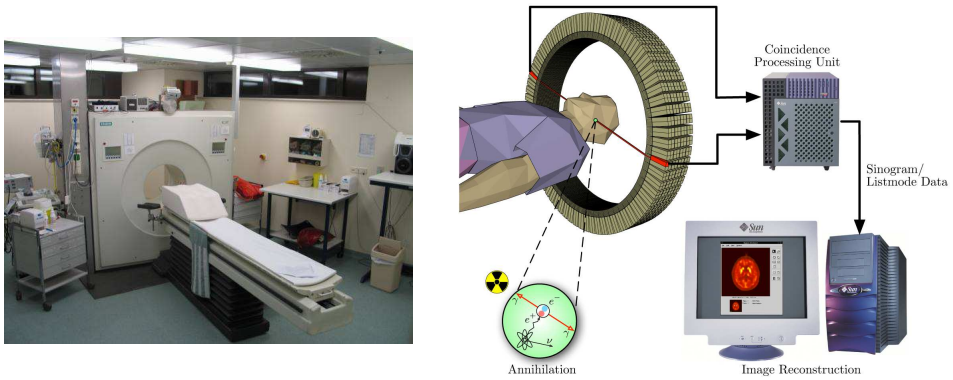


Figure 1.4: Illustration of positron emission tomography (PET). left: PET scanner, right: Visualization: From data to reconstruction.

modality can be used for instance to detect tumors, locate areas of the heart affected by coronary artery disease and identify brain regions influenced by drugs. Therefore,

PET is categorized as a functional imaging method and differs from methods such as X-ray computed tomography (CT) that measures attenuation related to density integrals on rays (non-functional). The data acquisition in PET is based on weak radioactively marked pharmaceuticals, so-called tracers, which are injected into the blood circulation. Used markers are suitable radio-isotopes which decay by emitting positrons annihilating almost immediately with electrons in the body of a patient. The resulting emissions of photons are collected by a detector system surrounding the patient, see Figure 1.4. Due to the radioactive decay, measured data can be modeled as an inhomogeneous Poisson process with a mean given by the X-ray transform of the spatial tracer distribution. The X-ray transform maps a function on $\mathbb{R}^d$ into the set of its line integrals. More precisely, if $\theta \in S^{d-1}$ and $x \in \Omega$, the operator describing our inverse problem is an integral of u over the straight line through x with direction θ,

$$(Ku)(\theta, x) \;=\; \int_{\mathbb{R}} u(x \,+\, t\,\theta)\; dt \;, \qquad x \,+\, t\,\theta \;\subseteq\; \Omega \;.$$

Up to notation, in the 2D case the X-ray transform is equivalent to the more popular Radon transform. To summarize, the image reconstruction task in PET is to generate density images of the inside of a patient from count data (sinograms) collected at the detectors.

For further examples of ill-posed operator equations we refer to [65, Chapter 1]. One difficulty of solving inverse problems in imaging is that often only noisy data are available, and hence the reconstruction of images gets unstable due to ill-posedness. In order to accomplish that, one can incorporate a priori knowledge about the solution into the reconstruction process. This means we use so-called regularization methods. To obtain variational methods, which are very common in imaging and image processing, one can use a Bayesian approach for computing solutions of the operator equations above in the presence of stochastic effects such as noise.

In a Bayesian approach one computes the MAP (maximum a-posteriori probability) estimate by

$$u_{MAP} \;=\; \arg\min_{u} \left(-\log p(u|f)\right) \;,$$

where $p(u|f)$ denotes the probability of observing u under given data f. By incorporating knowledge about the noise type of the data f and by incorporating a priori known smoothness properties of the solution u, one can derive variational problems of the form

$$\min_{u \in \mathcal{W}(\Omega)} \left\{H_f(Ku - f) \;+\; \alpha\, J(u)\right\} \;,$$

where H_f denotes a general data fidelity dependent on the operator K and the given

data f and where J denotes a general regularization term penalizing deviations from the smoothness of a solution.

First part of the thesis: 3D static image reconstruction
In the first part of the thesis we will concentrate on 3D static image reconstruction models concerning non-standard noise models. We will use total variation (TV) regularization and Bregman distances for the applications presented above. In the case of Poisson noise and TV regularization we obtain the following variational problem

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \ a.e.}} \int_{\Sigma} (Ku - f \log Ku) \, d\mu \ + \ \alpha \, |u|_{BV(\Omega)} \ , \tag{1.1}$$

where $\alpha > 0$ denotes a regularization parameter and where the total variation seminorm is defined as

$$|u|_{BV(\Omega)} \ := \sup_{\substack{g \in C_0^\infty(\Omega, R^d) \\ ||g||_\infty \leq 1}} \int_{\Omega} u \ \nabla \cdot g \ .$$

In recent years extensions of the so-called EM algorithm respectively Richardson-Lucy algorithm received increasing attention for inverse problems with Poisson data. However, most algorithms for regularizations like TV produce images suffering from blurred edges due to approximations in the algorithms, and neither can guarantee positivity nor provide detailed convergence proofs.

The **first goal** of this part is to provide an accurate, fast and robust EM-TV algorithm for efficient noise removal and for computing cartoon reconstructions facilitating post-segmentation. The method will be reinterpreted as a modified forward-backward splitting (FBS) strategy known from convex optimization. We want to establish the well-posedness of the basic variational problem in Section 1.1 including existence, uniqueness and stability, and we want to show that our EM-TV method preserves the positivity of a solution. In addition, we are interested in proving the convergence of proposed algorithms. A damped variant of the EM-TV algorithm with modified time steps, will be the key step towards convergence.

Typically, TV-based reconstruction methods provide reconstructions suffering from contrast reduction. More precisely, it was shown for instance by Meyer in [124] or in [134] that the amount of contrast reduction of an eigenmode (e.g. a cylinder in $2D$) of TV can be expressed by the regularization parameter α. Hence, the **second goal** in this first part of the thesis is to propose extensions to EM-TV, based on Bregman iterations and primal and dual inverse scale space methods, in order to improve imaging results by simultaneous contrast enhancement. In this thesis Bregman distances will be used in several parts for different applications playing different roles, e.g. as an iterative regularization technique, as an analytical tool or

as a key to fast L^1-type algorithms. The Bregman distance for a convex functional J is defined as follows

$$D_J^p(u,v) \;:=\; J(u) - J(v) - \langle p, u - v \rangle \,, \tag{1.2}$$

where $p \in \partial J(v)$ is a subgradient of J at v.

In the case of an L^2 data fidelity the iterative Bregman distance regularization concerning J reads as follows

$$\left\{ \begin{array}{rcl} u^{l+1} & = & \arg\min\limits_u \left\{ \dfrac{1}{2} \left\| Ku - f \right\|_{L^2(\Sigma)}^2 \;+\; \alpha \, D_J^{p^l}(u, u^l) \right\} \\ p^{l+1} & = & p^l \;-\; \dfrac{1}{\alpha} K^*(Ku^{l+1} - f) \;, \end{array} \right.$$

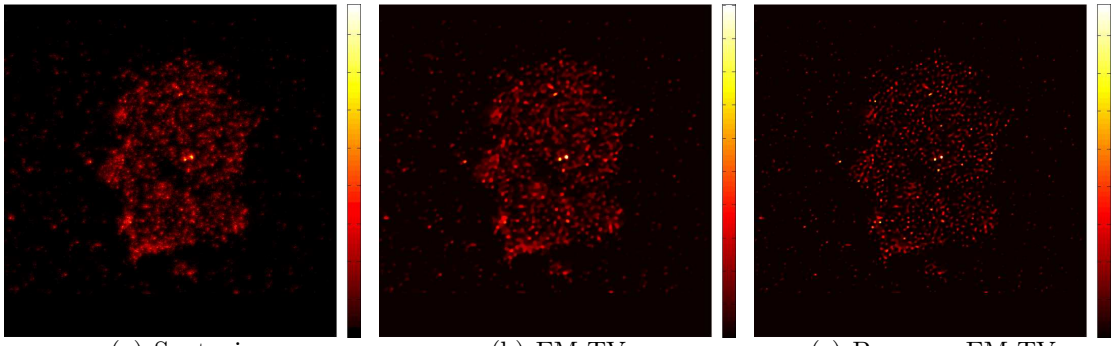
where the subgradient $p^l \in \partial J(u^l)$ can be interpreted as a residual or noise function. Via substitutions and by shifting terms to the quadratic data fidelity it is possible to simplify this problem to the following equivalent scheme

$$\left\{ \begin{array}{rcl} u^{l+1} & = & \arg\min\limits_u \left\{ \dfrac{1}{2} \left\| Ku - f^l \right\|_{L^2(\Sigma)}^2 \;+\; \alpha \, J(u) \right\} \\ f^{l+1} & = & f^l \;+\; f - Ku^{l+1} \;. \end{array} \right.$$

By adding back residuals, this method simultaneously enhances the contrast of reconstructions. However, for general data fidelities H_f, particularly those of non-L^2 type (i.e. not related to additive Gaussian noise), the question arises:

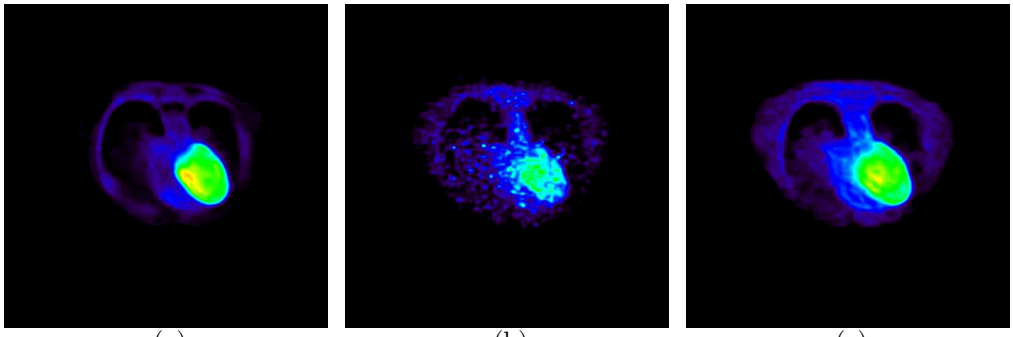
Is it always possible to find a Bregman regularization strategy that can be realized by a simple shift in the data fidelity?

In other words, the **third goal** of this part is to find a Bregman iteration for general data fidelities that simultaneously enhances contrast and is easy to implement. This issue yields the general concept of dual Bregman regularization strategies and makes it possible to obtain error estimates respectively convergence rates. We will illustrate the performance of our techniques and analytical concepts by 2D and 3D synthetic and real-world results in optical nanoscopy such as in Figure 1.5 or in PET such as in Figure 1.6. In Figure 1.5 you can see a significant reconstruction improvement if you compare the original noisy and blurred STED data with the EM-TV and Bregman-EM-TV reconstructions.



(a) Syntaxin (b) EM-TV (c) Bregman-EM-TV

Figure 1.5: 2D experimental data: (a) Protein Syntaxin in cell membrane, fixed mamalian (PC12) cell; image size: 1000 x 1000; (b) EM-TV reconstruction; and (c) 3rd iterate u^3 of the Bregman-EM-TV algorithm.



(a) (b) (c)

Figure 1.6: PET results of the heart: (a) 3D PET, 20 min ("ground truth"); (b) 3D EM, 5 sec; (c) 3D EM-TV, 5 sec

Second part: Joint 4D Image Reconstruction and Optimal Transport

In the second part of this thesis we will extend 3D image reconstruction to 4D image sequence reconstruction. Instead of a static inverse problem, $Ku = f$, we now have to solve inverse problems in space and time, i.e.

$$\tilde{K}(\rho(x,t)) \;=\; f(x,t), \qquad (x,t) \in \Omega \times [0,T] \subset \mathbb{R}^3 \times \mathbb{R} \;.$$

For fixed time steps $t = t_0$, intuitively, one may take into account the application of static reconstruction techniques similarly to the first part of the thesis.

However, since we talk about time dependent inverse problems, we have to consider that "life is always in motion". In particular, in our applications in 4D medical imaging and video fluorescence microscopy we can observe natural motion effects. For example in positron emission tomography we have to consider natural patient

motion like breathing or beating of the heart. To be more precise, in clinical studies it has been found that the breathing motion of the diaphragm can range from 4 mm to 38 mm, cf. [179].

Similarly, in dynamic fluorescence microscopy (e.g. dynamic STED) we can think of reconstructing videos in live cell imaging. Here we have motion effects for instance due to the migration of single cells or proteins. In optical nanoscopy time dependent reconstruction algorithms and tracking techniques are becoming more and more important. For example, recently, biophysicist achieved to observe and study living cells in nano-scale, cf. [95, 115]. For the reconstruction of 4D data in medical imaging or video microscopy standard methods as simple generalizations of 3D reconstruction algorithms can be applied, but usually they do not incorporate time dependent motion models. However, in dynamic tomography, e.g. in positron emission tomography (PET), motion is a well-known source of degradation of reconstruction results. Hence reconstruction models without incorporating motion models can cause significant blurring effects in resulting image sequences, especially at object boundaries, cf. Figure 7.1 (artifacts at the cardiac wall in the bottom image).

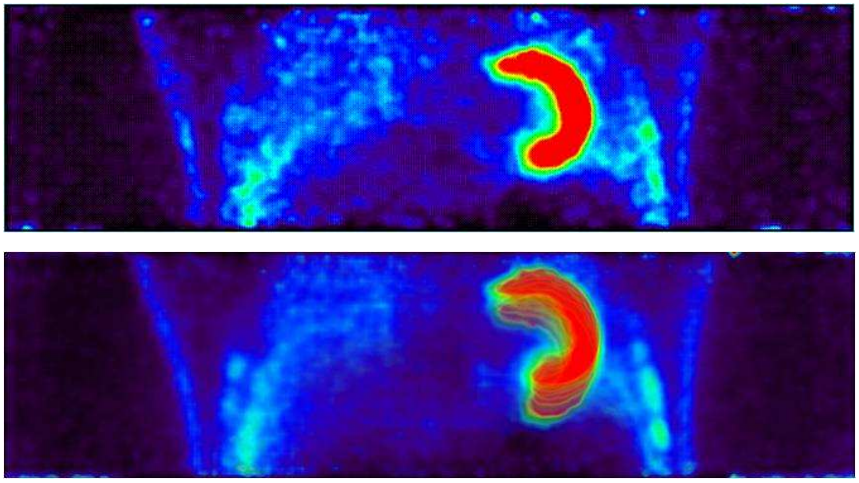


Figure 1.7: Comparison of a PET sequence reconstruction with and without motion modeling; Blurring effects due to motion artifacts in the bottom image. The data is from the European Institute for Molecular Imaging (EIMI).

Here we can see undesired doubling effects. The blurring effects on the PET images usually are proportional to the magnitude of the motion. In the literature it is well-known that these motion artifacts can cause significant errors in a later quantification. For example motion artifacts can cause a wrong staging of tumors, cf. [135, 66], or they can cause incorrect uptake values, cf. [132], and it can also

happen that small tumors may remain undetected [137]. Therefore, the combination of parameter estimation models or motion models and reconstruction techniques is an interesting recently growing research area, cf. [123, 144, 80, 54, 147, 181].

As a consequence, our **first goal** for the second part of the thesis is to study and characterize various models for motion estimation and optimal transport.

In computer vision optical flow estimation deals with the computation of visual motion information in image sequences. For a given video $\rho(x,t)$ a common assumption for estimating the optical flow is brightness constancy,

$$\rho(x+\mathbf{v}, t+\delta t) \stackrel{!}{=} \rho(x,t) \ ,$$

i.e. following the desired velocity field $\mathbf{v}$, the pixel intensity should not change. By using a Taylor expansion this assumption implies the following PDE, the so-called optical flow constraint (OFC),

$$\partial_t \rho + \mathbf{v} \ \nabla_\Omega \rho = 0 \ .$$

Hence, a standard model for optical flow estimation reads as follows,

$$\min_{\mathbf{v}} \ \|\partial_t \rho + \mathbf{v} \ \nabla_\Omega \rho\|_2^2 \ + \ \alpha \ J(u) \ , \tag{1.3}$$

where the first term, the data fidelity term, penalizes deviations from the OFC and where $J(u)$ denotes a regularization term penalizing deviations from a certain smoothness of the flow field.

For dynamic tomography data of the heart we basically have two types of motion. On the one hand motion due to the respiratory displacement and on the other hand motion due to the cardiac contraction. We will point out that the optical flow concept only takes into account incompressible flows, whereas mass conservation is more general and also takes into account compressible flows which will be useful e.g. for cardiac contraction. In other words, mass conservation allows significant density changes due to time, whereas the mass is preserved. Hence, we pass over to continuity equations and optimal transport.

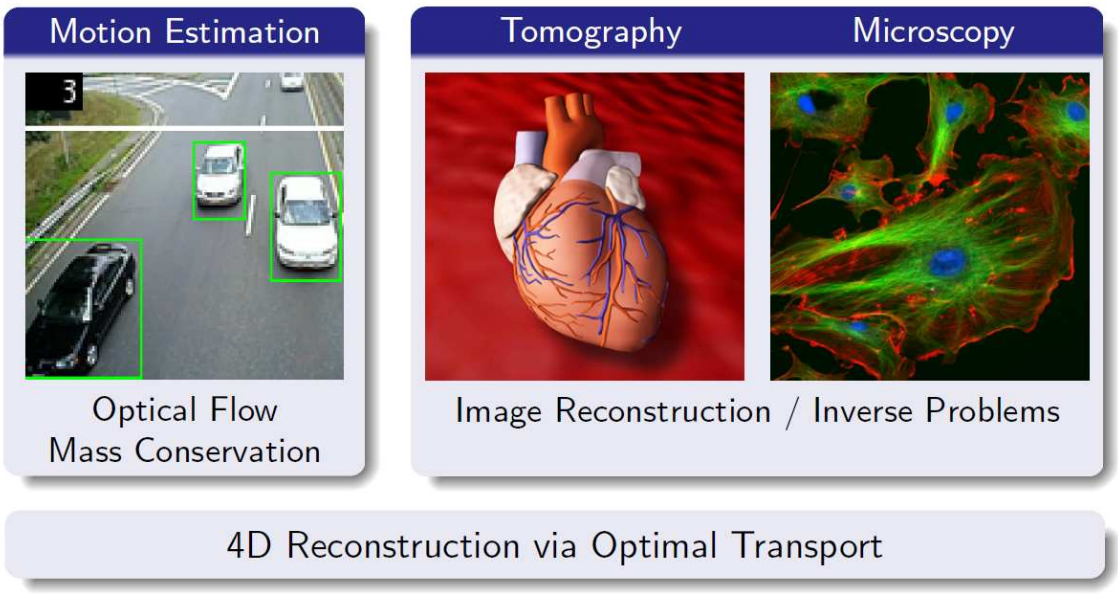
In Chapter 9 we will give an introduction to basic concepts of continuum mechanics. Based on this, we will study mass conservation and optimal transport further in detail. Particularly with regard to our joint 4D reconstruction model we will concentrate on the following model for optimal transport with mass conservation

$$\inf_{\rho,\mathbf{v}} \ T \int_0^T \int_\Omega \rho(x,t) \ |\mathbf{v}(x,t)|^2 \ dx \ dt$$

$$\begin{aligned} \text{subject to} \qquad \partial_t \rho + \nabla_\Omega \cdot (\rho\, \mathbf{v}) &= 0 \qquad \text{(mass conservation)} \\ \rho(\cdot, 0) &= \rho_0 \\ \rho(\cdot, T) &= \rho_T \ , \end{aligned}$$

which was introduced by Benamou and Brenier in [17, 18]. The main idea of this constrained optimization problem is to find an optimal transport "plan" $\mathbf{v}$, to move a given start density ρ_0 to a given end density ρ_T (the original problem of Monge [128] was to move a heap of sand into a hole of the same size). The continuity equation is the transport equation and the integral represents the transport cost.

Our **final goal** in the thesis is to combine the concepts of motion estimation and optimal transport with the reconstruction ideas of the first part of the thesis to build a joint 4D model for simultaneous image reconstruction, total variation regularization and optimal transport (including mass conservation), see Figure 7.1.



Our new 4D model for joint image reconstruction, spatio-temporal regularization and optimal transport, reads as follows

Model 1.1.1. *General 4D Reconstruction & Optimal Transport*

$$\begin{aligned} \min_{\rho, \boldsymbol{v}} \ & \int_0^T H_{f(\cdot,t)}\left(K\rho(\cdot,t)\right)\, dt \ + \ \alpha \int_0^T J(\rho(\cdot,t))\, dt \ + \ \frac{\beta}{2} \int_0^T \int_\Omega \rho\, |\boldsymbol{v}|^2\, dx\, dt \qquad (1.4) \\ s.t. \ & \partial_t\, \rho \ + \ \nabla_\Omega \cdot (\rho\, \boldsymbol{v}) \ = \ 0 \qquad \text{in } Q := \Omega \times [0, T] \\ & \rho \geq 0 \ . \end{aligned}$$

On the one hand, this model computes a sequence of reconstructed images ρ from the sequence of given data f in the sense of inverse problems while regarding an optimal transport scheme. On the other hand, the model simultaneously estimates the motion field $\mathbf{v}$ in the sequence, which can be used later on, e.g. for registration or tracking purposes. The first term of the objective functional is a data fidelity term H_f in space and time, which depends on the given image sequence f, on a compact operator K and on the desired density sequence ρ. The compact operator K is assumed to have the same structure in all time steps t. This makes sense for various applications since detector systems like microscopes or tomographs will not change their properties during time. In analogy to static reconstruction, different data fidelities will result from different noise models using Bayesian theory and MAP estimation. The second term of the objective functional is a time dependent regularization term. J can for example be a quadratic regularization term or the total variation. The third term and the constraint are related to the optimal transport scheme with mass conservation we introduced above. The main difference to optimal transport is the fact that we do not have a given start or end density. Initial values for the transport equation automatically arise from the given data of the inverse problem. In addition, we need to add a positivity constraint to the model since we consider density images.

For this general formulation we will study different variants including TV regularization in space and time. We will provide a detailed analysis including existence and uniqueness proofs. To overcome the large amount of data we will propose two types of numerical realizations based on preconditioning and splitting techniques to facilitate parallelization and efficiency. On the one hand we will present a preconditioned Newton-SQP method with integrated line-search. On the other hand, for the case of TV, we will present a special splitting technique based on inexact Uzawa techniques that is highly parallelized and where each of the resulting sub-steps in the algorithm is an efficient shrinkage or an efficient DCT inversion, which can additionally be parallelized on GPUs.

To summarize, in this thesis we will provide several models in 3D and 4D image reconstruction including motion estimation, a detailed analysis for different models, as well as a wide range of numerical realizations and results in real applications.

1.2 Organization of the Work

The thesis is organized as follows (also see Figure 1.2). The following three chapters, Chapter 2 - Chapter 4, serve as a general introduction of mathematical tools,

concepts and algorithms. The main ideas and techniques of these chapters will be applied in several sections throughout this thesis, see the left hand side of Figure 1.2.

In Chapter 2 we introduce variational methods for inverse problems and image processing tasks via the Bayesian model. In particular, we discuss different function spaces for image processing, e.g. the space of functions with bounded total variation BV, and we introduce the total variation (TV). Subsequently, we consider basic definitions and theorems in differentiability and optimality for a later analysis of models and algorithms.

Chapter 3 is dedicated to convex analysis and Bregman distances, see (1.2). We introduce concepts of subdifferential calculus and Legendre-Fenchel duality first. Then, we pass over to Bregman distances and the "role", the distance plays in this thesis. We study standard primal and a new dual Bregman iteration for general variational problems. The latter iteration has very nice properties. For example it enhances contrast in the sense of inverse scale space methods, and we can proof the well-posedness as well as error estimates. In the case of L^2 data fidelities this dual Bregman method coincides with the primal Bregman iteration, see (1.3). In a section on error forgetting we discuss recent ideas that may explain the power of Bregman iterations. The primal and dual Bregman methods will be applied to reconstruction problems with Poisson noise and TV in Chapter 5.

In Chapter 4 we study splitting methods in convex optimization. The main idea of these methods is to decouple a variational or a constraint optimization problem, such that resulting algorithms consist of cost-efficient sub-steps, which can be handled and adapted easily. We start with the introduction of general saddle point problems, because most splitting methods in this context can be characterized as primal and dual multi-step methods. Continuing with the general class of inexact Uzawa algorithms and with augmented Lagrangian methods, we aim preconditioning and build a general basis for the "zoo" of specific splitting techniques known in literature. Without making a claim to be complete we will introduce the main splitting techniques used in this thesis. More precisely, in Chapter 6 we will use forward-backward splitting for solving 3D TV reconstruction problems such as the one in (1.1). With the Split Bregman technique (resp. DRS or ADMM) we are going to establish a new algorithm for optical flow computation with TV regularization. In Chapter 10 we will make extensive use of inexact Uzawa strategies, respectively Bregmanized Operator Splitting (BOS), with preconditioning.

In Chapter 5 we will study 3D TV reconstruction problems in the case of Poisson noise as introduced above. Furthermore, we are going to present extensions to simultaneous contrast enhancement via Bregman iterations, and a detailed analysis of models and the proposed EM-TV algorithm.

In Chapter 6 the methods introduced in the previous chapter will be applied to 2D and 3D reconstruction problems in optical nanoscopy and positron emission tomography.
Chapter 7 serves as a "bridge" between 3D static image reconstruction and 4D time dependent imaging, providing an overview on the subsequent chapters.
In Chapter 8 we will discuss different motion models based on optical flow, such as the one in (1.3), further in detail. We will characterize different data fidelities and different smoothing terms for the flow field. In particular, we will combine optical flow techniques with total variation regularization and splitting techniques, and we will present results in high resolution computed tomography (CT) and tracking applications.
In Chapter 9 we will give an introduction to basic concepts of continuum mechanics. Based on this, we will study mass conservation and optimal transport problems as introduced above.
Chapter 10 is dedicated to our joint 4D image reconstruction model with optimal transport we introduced in Model 1.1.1.

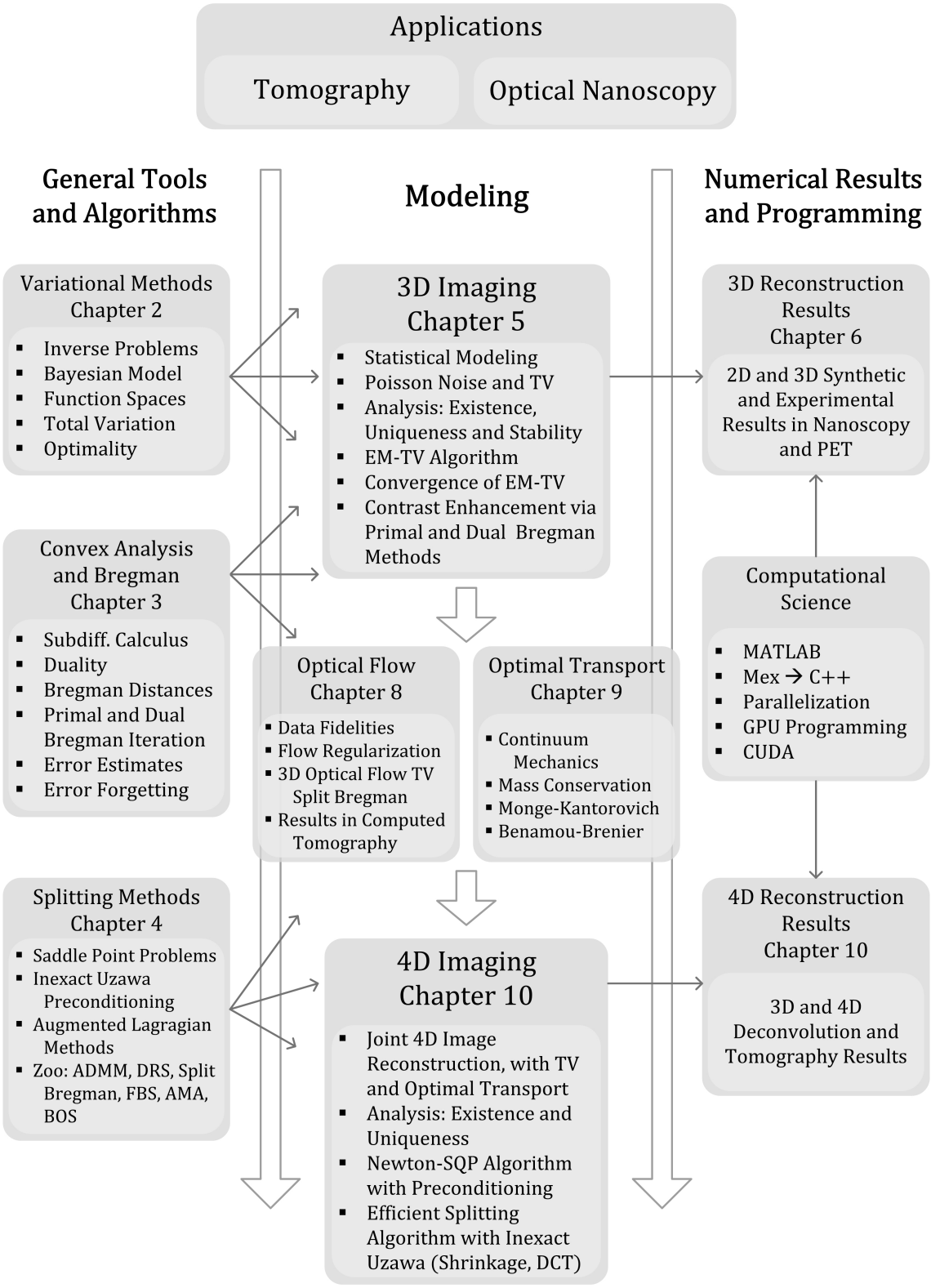


Figure 1.8: Organization of the thesis.

2 Variational Methods

2.1 Motivation: Inverse Problems

First of all, in order to derive general variational models for image and video processing tasks we need to define the terms image and image sequence in a continuous setting.

Definition 2.1.1 (Image)**.** *Let $\Omega \subset \mathbb{R}^d, d \in \mathbb{N}$ be an image domain. A function $u : \Omega \to \mathbb{R}$ is called a d-dimensional image if the following conditions are fulfilled,*

- *u has a compact support, if Ω is not bounded.*
- $0 \leq u(x) < \infty, \quad \forall x \in \Omega$ *(Intensity boundedness)*
- $\int_\Omega u(x)dx \leq \infty$ *(Energy boundedness) .*

If u is not only a one-channeled gray value image but also a color image, the definition extends for each channel in the straight forward way.

The definition of an image refers to a light intensity function u where each spatial point x is mapped to a certain gray value or color. For example in biomedical imaging we mainly think of density images. We denote the boundary of Ω by $\partial\Omega$.

In dynamic image processing applications we work with videos, i.e. a sequence of images.

Definition 2.1.2 (Image sequence)**.** *Let $Q = \Omega \times [0, T]$ be a space-time cylinder. A function $\rho : Q \to \mathbb{R}$ is called a d-dimensional image sequence if it is a d-dimensional image for every fixed $t \in [0, T]$.*

Since time dependence and motion plays an important role in this work, we will use u to denote static images and ρ for image sequences. Image reconstruction is a fundamental problem in many fields of applied sciences, e.g. nanoscopic imaging, medical imaging or astronomy. Fluorescence microscopy for example is an important imaging technique for the investigation of biological (live-) cells, down to nano-scale. In this case image reconstruction arises in form of deconvolution problems. Undesirable blurring effects can be ascribed to a diffraction of light.

Mathematically, image reconstruction in such applications can often be formulated as the computation of a function $\tilde{u} \in \mathcal{U}(\Omega)$ from the operator equation

$$K\tilde{u} \;=\; g \;. \tag{2.1}$$

Typically, this is a Fredholm integral equation of the first kind with given *exact data* $g \in \mathcal{V}(\Sigma)$ and the desired exact solution $\tilde{u}$.

Here, K denotes a linear and compact operator $K : \mathcal{U}(\Omega) \to \mathcal{V}(\Sigma)$ and $\mathcal{U}(\Omega)$ as well as $\mathcal{V}(\Sigma)$ are Banach spaces of functions on bounded and compact sets Ω respectively Σ. In the case of nanoscopic imaging $\tilde{u}$ is a convolution operator

$$(Ku)(x) \;=\; (k * u)(x) \;=\; \int_{\Omega} k(x-y)u(y)\,dy \;,$$

where k is a convolution kernel, describing the blurring effects caused by a nanoscopic apparatus.

Since K cannot be inverted continuously (due to the compactness of the forward operator), most inverse problems are ill-posed. Furthermore, in real-life applications the exact data g are usually not available, but a noisy version f instead. Hence, we need to compute approximations to the ill-posed problem 2.1, i.e.

$$Ku = f \tag{2.2}$$

with $u \in \mathcal{U}(\Omega)$ and $f \in \mathcal{V}(\Sigma)$.

2.2 Bayesian Modeling

In order to obtain general variational models including different data fidelities and different regularization terms we make use of the Bayesian approach for computing solutions of the operator equation in 2.2 in the presence of stochastic effects such as noise. In a classical log-likelihood estimation technique one computes a solution by minimizing the negative log-likelihood of observing f under u, i.e.

$$u_{ML} = \arg\min_{u} \left(-\log p(f|u)\right) \;,$$

where $p(f|u)$ denotes an appropriate probability density for observing f given u. This can usually be identified with the probability density of the noise, e.g. in the case of additive Gaussian noise

$$f = Ku + \eta \ ,$$

where η denotes a stochastic perturbation (noise), we obtain

$$p(u|f) = p_\eta(f - Ku) \ . \tag{2.3}$$

A frequently used way to realize the latter is the Bayesian model, whose aim is the computation of an estimate u of the unknown object by maximizing the a-posteriori probability density $p(u|f)$ with measurements f. In other words, in a Bayesian approach one computes the MAP (maximum a-posteriori probability) estimate by

$$u_{MAP} = \arg\min_u \left(-\log p(u|f)\right) \ . \tag{2.4}$$

The posterior probability density is given according to Bayes formula

$$p(u|f) = \frac{p(f|u)\, p(u)}{p(f)} \ , \tag{2.5}$$

where $p(u)$ is the prior probability for u and where $p(f)$ denotes the prior probability for f. Since $p(f)$ does only contribute a constant term in the minimization of 2.4 we can write the MAP estimate similarly to the ML estimate above as

$$u_{MAP} = \arg\min_u \left(-\log p(f|u) - \log p(u)\right) \ .$$

The MAP approach has the advantage that it allows to incorporate additional information about u via the prior probability density $p(u)$ into the reconstruction process. It acts as a penalty or regularization functional. The most frequently used prior densities are Gibbs functions

$$p(u) \sim e^{-\alpha\, J(u)} \ , \tag{2.6}$$

where α is a positive parameter and J a convex regularization energy $J : \mathcal{W}(\Omega) \to \mathbb{R} \cup \{\infty\}$ [78, 79].

Since stochastic modeling is often done in discrete terms based on the modeling of random variables, we introduce a semi-discrete, linear and compact operator

$$\overline{K} : \mathcal{U}(\Omega) \to \mathcal{D}(\Sigma) \, , \tag{2.7}$$

with a finite-dimensional range $\mathcal{D}(\Sigma)$, to be able to derive corresponding continuum models. Typical models for the probability density $p(f|u)$ in (2.5) are data with

Gaussian noise, Poisson-distributed data or Γ-distributed data with multiplicative noise, i.e.

$$\begin{aligned} p(f|u) &\sim e^{-\frac{1}{2\sigma^2}\|Ku-f\|^2_{L^2(\Sigma)}} && \text{(Gaussian)} \\ p(f|u) &= \prod_i \frac{(\overline{K}u)_i^{f_i}}{f_i!} e^{-(\overline{K}u)_i} && \text{(Poisson)} \\ p(f|u) &= \prod_i \frac{n^n}{(\overline{K}u)_i^n \, \Gamma(n)} f_i^{n-1} e^{-n\frac{f_i}{(\overline{K}u)_i}} && \text{(Gamma)} \ , \end{aligned} \tag{2.8}$$

where $\bar{K}$ is a semi-discrete operator derived from local mapping $\bar{K} = \int K$.

Most works deal with the case of additive Gaussian distributed noise so far. However, in real-life there are several applications, in which different types of noise are of a certain interest, such as Positron Emission Tomography (PET), Microscopy, CCD cameras, or radar. For instance, in addition to fluorescence microscopy, Poisson noise appears also in PET in medical imaging. Other non-Gaussian noise models are salt and pepper noise or the different types of multiplicative noise, for example appearing in Synthetic Aperture Radar (SAR) imaging to reduce speckle noise. In [6], Aubert and Aujol assumed η in $f = (Ku)\eta$ to follow a gamma law with mean one and derived the conditional probability $p(f|u)$ above. For such cases different variational models (fidelities related to the log-likelihood of the noise distribution) can be derived in the framework of MAP estimation, which need different analysis than in the case of Gaussian distributed noise.

In the canonical case of additive Gaussian noise (see (2.8)), the minimization of the negative log likelihood function (2.5) leads to classical Tikhonov regularization [22] based on minimizing a functional of the form

$$\min_{u \geq 0} \left\{ \frac{1}{2} \|Ku - f\|^2_{L^2(\Sigma)} \ + \ \alpha \, J(u) \right\} \ . \tag{2.9}$$

The first term, the so-called data-fidelity term, penalizes the deviation from equality in (2.1) whereas J is a regularization term as in (2.6). If we choose $K = Id$ and the total variation (TV) regularization technique $J(u) := |u|_{BV}$, we obtain the well-known ROF-model [157] for image denoising. The additional positivity constraint is necessary in typical applications as the unknown represents a density image.

In nanoscopic imaging measured data are stochastic and pointwise, more precisely, the data are called "photon counts". This property refers to laser scanning techniques in fluorescence microscopy. Consequently, the random variables of measured data are not Gaussian- but Poisson-distributed (see (2.8)), with expected value given by $(\overline{K}u)_i$. Hence an MAP estimation via the negative log likelihood function (2.5) leads

to the following variational problem [22]

$$\min_{u\geq 0} \left\{ \int_{\Sigma} (Ku - f \log Ku)\, d\mu \;+\; \alpha\, J(u) \right\} . \tag{2.10}$$

Up to additive terms independent of u, the data-fidelity term is the so-called Kullback-Leibler functional (also known as cross entropy or I-divergence) between the two probability measures f and Ku. A particular complication of (2.10) compared to (2.9) is the strong nonlinearity in the data fidelity term and resulting issues in the computation of minimizers.

The specific choice of the regularization functional J in (2.10) is important for the way a-priori information about the expected solution is incorporated into the reconstruction process. Smooth, in particular quadratic regularizations have attracted most attention in the past, mainly due to the simplicity in analysis and computation. However, such regularization approaches always lead to blurring of the reconstructions, in particular they cannot yield reconstructions with sharp edges.

Recently, singular regularization energies, in particular those of ℓ^1 or L^1-type, have attracted strong attention. Throughout this work, the total variation (TV) regularization functional will play a fundamental role. TV regularization has been derived as a denoising technique in [157] and has been generalized to various other imaging tasks subsequently. The exact definition of TV [1], used in this work, is

$$J(u) \;:=\; |u|_{BV} \;=\; \sup_{\substack{g\in C_0^\infty(\Omega;\mathbb{R}^d)\\ ||g||_\infty\leq 1}} \int_\Omega u \,\mathrm{div} g \;, \tag{2.11}$$

which is formally (true if u is sufficiently regular) $|u|_{BV} = \int_\Omega |\nabla u|\, dx$. The motivation for using TV is the effective suppression of noise and the realization of almost homogeneous regions with sharp edges. These features are attractive for nanoscopic imaging if the goal is to identify object shapes that are separated by sharp edges and shall be analyzed quantitatively.

2.3 Variational Calculus

The minimization of functionals as in (2.9) or (2.10) is related to the calculus of variations. A general form of optimization problems in image reconstruction and inverse problem is given by

$$\min_{u\in\mathcal{U}(\Omega)} \{\; H_f(Ku) \;+\; J(u)\} \;. \tag{2.12}$$

Taking a closer look at this, the question arises, over which class of functions $\mathcal{U}(\Omega)$ we have to minimize. In other words, we have to think about adequate regularization terms J, respectively function spaces for our imaging problems. The choice of a function space directly influences reconstruction results and the analysis of such problems, in particular the existence and uniqueness of minimizing elements.

2.3.1 Function Spaces and Total Variation

In this section we will introduce basic function spaces with special regard to mathematical image processing. Let $\Omega \subset \mathbb{R}^d$ be a bounded and open domain. Motivated by analytical investigations in image processing, Ω is usually assumed to be a Lipschitz space. For instance, we can take a closed rectangle, where the interior represents a strict Lipschitz space.

Spaces of continuously differentiable functions $\mathcal{C}^k$
$\mathcal{C}^k(\Omega, \mathbb{R}^n)$ for $1 \leq k < \infty$ is a space that contains all k-times continuously differentiable functions with values in $\mathbb{R}^n$. With $\mathcal{C}^\infty(\Omega, \mathbb{R}^n)$ we denote the space of infinitely differentiable functions with values in $\mathbb{R}^n$. We write $\mathcal{C}^k(\Omega)$ for $n = 1$. On the interior of Ω, the functions in $\mathcal{C}^k(\bar{\Omega}, \mathbb{R}^n)$ are uniformly continuous.

Lebesgue-Spaces L^p

Definition 2.3.1 ($\mathcal{L}^p$-Spaces)**.** *Let $\Omega \subset \mathbb{R}^m$ be an open set and let $1 \leq p < \infty$. With $\mathcal{L}^p(\Omega)$ we denote the space of all measurable functions $u : \Omega \to \mathbb{R}$ with the property*

$$\int_\Omega |u(x)|^p \, dx < \infty \ . \tag{2.13}$$

If $u : \Omega \to \mathbb{R}^n$ with $u^i \in \mathcal{L}^p(\Omega)$ holds for all $i = 1, ..., n$, we write $u \in \mathcal{L}^p(\Omega, \mathbb{R}^n)$.

For the vector space $\mathcal{L}^p$ the mapping

$$||u||_p \ := \ \left(\int_\Omega ||u(x)||^p \, dx \right)^{\frac{1}{p}} \qquad , u : \Omega \to \mathbb{R}^n \tag{2.14}$$

is only a seminorm, because we do not have definiteness. The triangle inequality for this seminorm is called Minkowski inequality and can be proved via the Hölder inequality. The completeness of the $\mathcal{L}^p$ spaces equipped with the seminorm is given by Riesz. From the space $\mathcal{L}^p$ with the seminorm we obtain a normalized space

$$L^p \ := \ \mathcal{L}^p \ / \sim \tag{2.15}$$

as a factor space. The following equivalence relation implies a unique zero element:

$$u \sim \tilde{u} \quad :\Longleftrightarrow \quad \lambda^m(\{x \in \Omega \mid u(x) \neq \tilde{u}(x)\}) = 0 \ , \tag{2.16}$$

where λ^m denotes the Lebesgue measure. Hence, functions are equivalent, i.e. they are in the same equivalence class, if they only differ in a point measure. A norm on L^p is defined as

$$|| \, [u] \, ||_p \; := \; ||u||_p \; . \tag{2.17}$$

Thus, an L^p-function can be interpreted as a whole equivalence class. For $p = \infty$ one can define an L^p-space as well. It is the space of measurable and essentially (almost everywhere) bounded functions $u : \Omega \to \mathbb{R}^n$. Thereby, the norm is given by the essential supremum

$$||u||_\infty \; := \; \text{ess} \sup_{x \in \Omega} ||u(x)|| \; = \; \inf_{\substack{\mathcal{N} \in \mathcal{B}(\Omega) \\ \lambda^m(\mathcal{N}) = 0}} \sup_{x \in \Omega \setminus \mathcal{N}} ||u(x)|| \; , \tag{2.18}$$

which is defined via the supremum of functions without regarding point measures. $\mathcal{B}(\Omega)$ denotes a σ-algebra of Borel sets of Ω. The spaces $L^p(\Omega, \mathbb{R}^n), \; 1 \leq p \leq \infty$, are Banach spaces equipped with the described norms. With the norm induced by the following scalar product,

$$\langle u, \tilde{u} \rangle_{L^2(\Omega, \mathbb{R}^n)} \; := \; \int_\Omega \langle u(x), \tilde{u}(x) \rangle \; dx \; , \tag{2.19}$$

the space $L^2(\Omega, \mathbb{R}^n)$ is a Hilbert space. Since L^p spaces are defined via Lebesgue integrals, they are called Lebesgue spaces. Two different continuous functions never lie in the same equivalence class. Hence, L^p spaces build a (natural) extension of continuous functions.

Sobolev Spaces $W^{k,p}$

Sobolev spaces form the basis of the solution theory of partial differential equations. For their introduction we need the concept of weak derivatives first. To define them, we need functions, which are zero on the boundary of Ω. We consider the space of test functions

$$\mathcal{C}_0^\infty(\Omega, \mathbb{R}^n) \; := \; \{ \; \varphi \in \mathcal{C}^\infty(\Omega, \mathbb{R}^n) \mid \text{supp}(\varphi) \subset \Omega \text{ is compact } \} \; , \tag{2.20}$$

the space of infinitely often differentiable functions with compact support on Ω. The support of φ is defined as the closed non-zero set of φ, i.e.

$$\text{supp}(\varphi) \; := \; \overline{\{x \in \Omega \mid \varphi(x) \neq 0\}} \; . \tag{2.21}$$

To introduce Sobolev spaces we need to relax the concept of derivatives, in order to ascribe derivatives to functions that are not differentiable in the "usual" sense. In particular, the partial integration can be guaranteed further on.

Definition 2.3.2 (Locally integrable functions)**.** *A function $u : \Omega \to \mathbb{R}^n$, which is Lebesgue integrable on every compact subset of Ω is called a locally integrable function. The set of all locally integrable functions in Ω is denoted by $L^1_{loc}(\Omega, \mathbb{R}^n)$.*

By considering multiindices we define the weak derivative as

Definition 2.3.3 (Weak derivative)**.** *Let $u \in L^1_{loc}(\Omega, \mathbb{R}^n)$ and let $\alpha \in \mathbb{N}_0^m$ be a multiindex. Then, if there exists a function $\omega \in L^1_{loc}(\Omega, \mathbb{R}^n)$ with*

$$\int_\Omega u(x) D^\alpha \varphi(s) dx = (-1)^{|\alpha|} \int_\Omega \omega(x)\varphi(x) dx \ , \qquad \forall\ \varphi \in \mathcal{C}_0^\infty(\Omega) \ , \tag{2.22}$$

we call ω a weak derivative of order $|\alpha|$ of u and denote the latter with $\omega = D^\alpha u$.

In other words, ω is a weak derivative of order α of u if it fulfills the partial integration at positions of the (strong) derivative $D^\alpha u$. Let us have a look at the following simple example:

Example 2.3.4 (Absolute value function)**.** *Let $u(x) = |u|$ in $\Omega = (-1, 1)$.*
Then u has the weak derivative

$$u'(x) = \omega(u) = \begin{cases} -1 & x \in (-1, 0) \\ 1 & x \in [0, 1) \end{cases} , \tag{2.23}$$

because for all $\varphi \in \mathcal{C}_0^\infty(-1, 1)$,

$$\begin{aligned} \int_{-1}^1 |x| \varphi'(x) dx &= \int_{-1}^0 (-x)\varphi'(x) dx + \int_0^1 x\varphi'(x) dx \\ &= -x\varphi(x)|_{-1}^0 - \int_{-1}^0 -1\varphi(x) dx + x\varphi(x)|_0^1 - \int_0^1 1\varphi(x) dx \\ &= -\int_{-1}^1 \omega(x)\varphi(x) dx \end{aligned} \tag{2.24}$$

holds. The value at zero can be chosen arbitrarily, since it forms a point measure.

A weak derivative of a function can exist, but does not have to. The space $L^1_{loc}(\Omega, \mathbb{R}^n)$ is not convenient, such that one is more interested in weak derivatives belonging to $L^p(\Omega, \mathbb{R}^n)$ spaces. This leads to the definition of Sobolev spaces:

Definition 2.3.5 (Sobolev Spaces)**.** *Let $1 \leq p < \infty$ and $k \in \mathbb{N}$. With $W^{k,p}(\Omega, \mathbb{R}^n)$ we denote the space of all functions $u \in L^p(\Omega, \mathbb{R}^n)$, for which the weak derivatives $D^\alpha u$ exist for all multiindices α with $|\alpha| \leq k$ and also belong to $L^p(\Omega, \mathbb{R}^n)$, i.e.*

$$W^{k,p}(\Omega, \mathbb{R}^n) \ := \ \{u \in L^p(\Omega, \mathbb{R}^n) \ : \ D^\alpha u \in L^p(\Omega, \mathbb{R}^n),\ |\alpha| \leq k\} \tag{2.25}$$

For functions $u \in W^{k,p}(\Omega, \mathbb{R}^n)$ the $W^{k,p}$-norm is defined as:

$$||u||_{W^{k,p}(\Omega,\mathbb{R}^n)} \ := \ \left(\sum_{|a| \leq k} ||D^\alpha u||^p_{L^p(\Omega,\mathbb{R}^n)} \ dx \right)^{\frac{1}{p}} . \tag{2.26}$$

Similarly we can introduce $W^{k,\infty}(\Omega, \mathbb{R}^n)$ *for* $p = \infty$ *equipped with the norm*

$$||u||_{W^{k,\infty}(\Omega,\mathbb{R}^n)} := \max_{|\alpha|\leq k} ||D^\alpha u||_{L^\infty(\Omega,\mathbb{R}^n)} . \tag{2.27}$$

The spaces $W^{k,p}(\Omega, \mathbb{R}^n)$ *are Banach spaces and are called Sobolev spaces.*

For $p = 2$ the norm is induced by the scalar product

$$\langle u, \tilde{u}\rangle_{W^{k,2}(\Omega,\mathbb{R}^n)} := \sum_{|\alpha|\leq k} \langle D^\alpha u, D^\alpha \tilde{u}\rangle_{L^2(\Omega,\mathbb{R}^n)} , \tag{2.28}$$

where $\langle \cdot , \cdot \rangle_{L^2(\Omega,\mathbb{R}^n)}$ denotes the scalar product (2.19) of the Hilbert space L^2. Thus $W^{k,2}(\Omega, \mathbb{R}^n)$ is a Hilbert space and of special interest. An additional and common notation for the latter is given by

$$H^k(\Omega, \mathbb{R}^n) := W^{k,2}(\Omega, \mathbb{R}^n), \quad \forall k \in \mathbb{N} . \tag{2.29}$$

The space $H^1(\Omega, \mathbb{R}^n)$ is commonly used in literature.

Space of functions of bounded total variation, BV

In this section we will generalize the term (weak) derivative further and we will introduce the so called distributional derivative. To simplify notation we will consider functions $u : \mathbb{R} \supset \Omega \to \mathbb{R}$ first. A distribution can be seen as a generalization of a function. More precisely, a distribution is a continuous, linear functional on a set of test functions. Thus

$$\begin{aligned} T : \mathcal{C}_0^\infty(\Omega) &\to \mathbb{R} \\ \varphi &\mapsto T(\varphi) = \langle T, \varphi\rangle \ . \end{aligned} \tag{2.30}$$

To understand distributions, we examine in which way they generalize functions. Let u be a piecewise continuous function and let $\varphi \in \Omega$ be a test function, such that $\int_\Omega u(x)\varphi(x)dx$ exists. Then we obtain a distribution T_u in the space of distributions $\mathcal{D}'(\Omega)$ with:

$$T_u(\varphi) = \langle T_u, \varphi\rangle = [u](\varphi) := \int_\Omega u(x)\varphi(x)dx . \tag{2.31}$$

As you can see, except for a point measure, u is uniquely determined by its distribution $T_u = [u]$. Distributions, which have been constructed in this way are called regular. The distributional derivative then is defined analog to the weak derivative from above:

$$D^\alpha T(\varphi) := (-1)^{|\alpha|} T(D^\alpha \varphi), \quad \forall \varphi \in \mathcal{C}_0^\infty(\Omega) . \tag{2.32}$$

The formula directly results from the application of partial integration, where the boundary terms drop out due to the compact support of the test functions φ. The

δ-distribution is an example of a non regular distribution. You can directly see, that there is no function in the ordinary sense representing the following formal properties:

$$\delta(x) := \begin{cases} 0 & x \neq 0 \\ \infty & x = 0, \end{cases} \qquad \text{with} \int_\Omega \delta(x)\, dx = 1 \ . \tag{2.33}$$

Accordingly, the Dirac-δ as a distribution is defined as follows:

$$\delta(\varphi) = \langle \delta, \varphi \rangle := \varphi(0) \ . \tag{2.34}$$

The idea of the distributional derivative can be illustrated nicely via the Heaviside

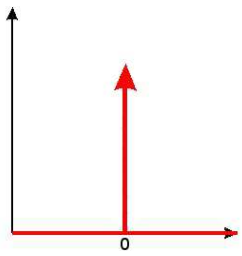


Figure 2.1: Dirac delta distribution with infinite momentum at 0

function Θ. The jump function is defined as

$$\Theta : \Omega \to \mathbb{R} \ \text{ with } \ \Theta(x) := \begin{cases} 0 & x \leq 0 \\ 1 & x > 0, \end{cases} \quad , \ \Omega = \mathbb{R}. \tag{2.35}$$

Considering the latter as a function it is neither continuous nor differentiable in zero. Considering it as distribution, we can formulate a distributional derivative. The latter is simply given by the δ distribution, because:

$$\begin{aligned} \Theta'(\varphi) &= -\Theta(\varphi') = -\int_0^\infty \varphi'(x)\, dx \qquad \forall \varphi \in \mathcal{C}_0^\infty(\Omega) \quad , \ \Omega = \mathbb{R} \\ &\overset{def.\ \varphi}{=} \varphi(0) = \delta(\varphi) \ . \end{aligned} \tag{2.36}$$

In this example we could see that passing from functions on to distributions, offers the definition of derivatives even at jump positions. With generalizations of that kind we can introduce the space of functions with bounded total variation, BV.

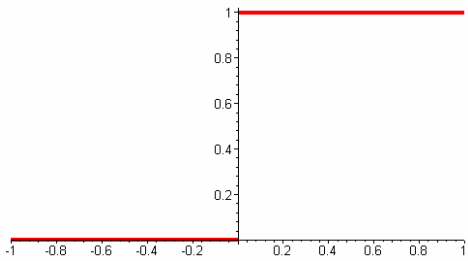


Figure 2.2: Heaviside function with jump at 0

Considering a BV function space in a model allows to compute solutions with sharp edges in images. In other words, we allow discontinuities and jumps in the intensities of a solution.

In imaging and image processing there are various task, where cartoon reconstructions, i.e. reconstructions with homogeneous regions and discontinuities are desired. The standard model for denoising with total variation is the Rudin, Osher, Fatemi (ROF) model [158]. Other applications for TV are deblurring, inpainting, segmentation models of the Mumford-Shah type or vectorial TV regularization in motion estimation. Besides the applicability of TV, the theory of BV-functions and the analysis of related models is an interesting area.

Using the example of a Heaviside function θ we could see that the distributional derivative never can be an L^p-function, but a measure on Ω. Provided, that u is smooth enough, we can define the total variation of u by

$$TV[u] \;:=\; \int_\Omega |\nabla u|\, dx \qquad\qquad , \; \Omega \subset \mathbb{R} \; , \tag{2.37}$$

first of all. If $TV(u) < \infty$ holds, we say u has bounded total variation. Therewith, we can provide a first definition of BV, the space of functions of bounded total variation,

$$BV(\Omega) \;:=\; \{u \in L^1(\Omega) \mid TV[u] < \infty\} \; . \tag{2.38}$$

Intuitively, a question arises: How smooth should u be, such that the definition makes sense? Although the Sobolev space $W^{1,1}(\Omega)$ seems to be a possible choice at first glance, the effectiveness of TV is based on the fact that we do not restrict to functions in $W^{1,1}(\Omega)$. By weakening the conditions for the function space $W^{1,1}(\Omega)$ we can attain functions with the desired properties described above. We will see that

$$W^{1,1}(\Omega) \subset BV(\Omega) \tag{2.39}$$

holds. In the following we will approach the BV-space from measure theory. A radon measure is a measure on the σ algebra of Borel sets $\mathcal{B}(\Omega)$, which is locally

finite and inner regular [64]. For example the non-negative measure μ on $\mathcal{B}(\Omega)$ with $\mu(K) < \infty$ for all compact spaces $K \subseteq \Omega$ is a Radon measure. To relate the terms bounded total variation and Radon measure, we introduce this topic via a simple total integral $F[\cdot]$. For distributions, e.g. the Dirac δ-distribution (2.34), we can write the total integral in a well-defined form

$$F[\delta] \;:=\; \int_{\Omega} \delta(x)\; dx = 1\;. \tag{2.40}$$

A set function is called signed measure μ, if $\mu(\emptyset) = 0$ holds, μ only takes one of the values $(-\infty), (+\infty)$ and μ is σ-additive. More general, a total integral can be defined for signed Radon measures $\mu \in \mathcal{M}(\Omega)$:

$$F[\mu] \;:=\; \int_{\Omega} 1\; d\mu = \mu(\Omega)\;. \tag{2.41}$$

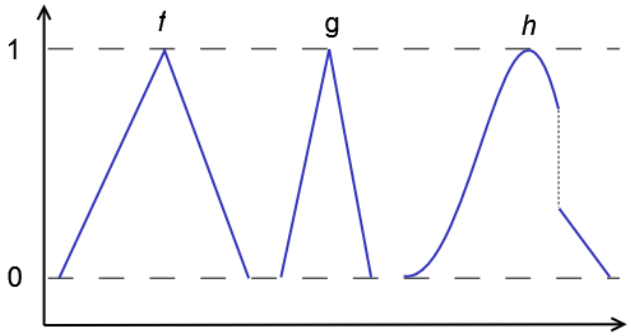
The δ-distribution from above can be interpreted as a bounded measure, more precisely, as a Radon measure. The space of signed Radon measures generalizes the space $L^1_{loc}(\Omega)$ of L^1 functions with locally compact support and actually represents the most general space, where the total integral $F[\cdot]$ is still well-defined. Now we want to transfer this ideas to the TV functional. For a signed Radon measure μ and its accumulated distribution we obtain

$$TV[\mu] \;=\; \int_{\mathbb{R}} 1\; |d\mu| \;=\; |\mu|(\mathbb{R})\;. \tag{2.42}$$

$|\mu|$ denotes the TV measure of μ, thus

$$|\mu| = \mu^+ + \mu^-,\; \text{for } \mu = \mu^+ - \mu^-\;. \tag{2.43}$$

Figure 2.3: 1D examples with $TV[f] = TV[g] = TV[ht] = 2$



With the latter concepts in mind we are able to redefine $TV[x]$ and the space $BV(\Omega)$. Let $U \subset \Omega$ be open and $u \in L^1(\Omega) = L^1(\Omega, \mathbb{R})$. Let

$$\int_U |D^* u| \;:=\; \sup_{\substack{g \in C_0^1(U) \\ ||g||_{L^\infty(\Omega)} \leq 1}} \int_U u\; \nabla \cdot g\; dx\;. \tag{2.44}$$

Hence

$$BV_{loc}(\Omega) \;:=\; \{u \in L^1(\Omega) \mid \int_{\bar{U}} |D^*u| < \infty \qquad \forall \bar{U} \subset \Omega \text{ compact } \} \tag{2.45}$$

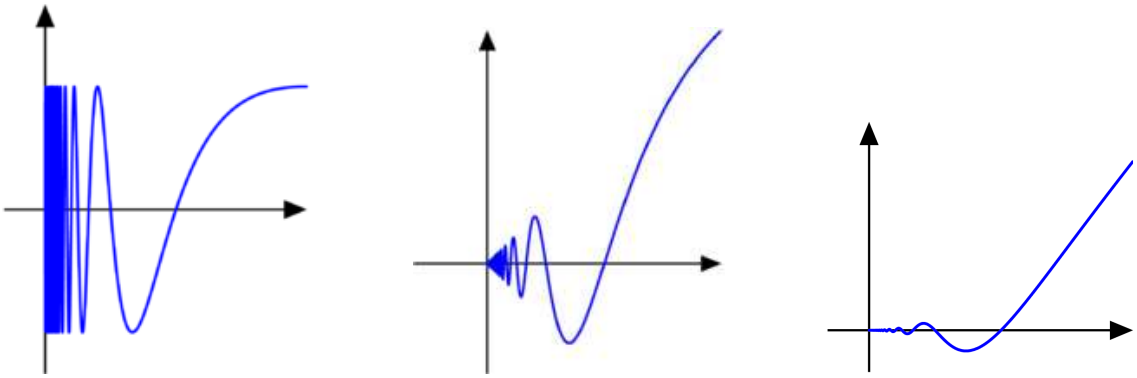
is the space of functions, which have locally bounded variation. Now we have to carry over this to an outer measure. For an arbitrary subset $E \subseteq \Omega$ we get

$$\int_E |D^*u| \;:=\; \inf_{U:\ E \subseteq U} \int_U |D^*u| \ . \tag{2.46}$$

Regarding measurable sets this yields a radon measure on Ω and we can define the total variation as

$$TV[u] \;:=\; |Du|(\Omega) \;=\; \int_\Omega |Du| \;=\; \sup_{\substack{g \in \mathcal{C}_0^1(\Omega) \\ ||g||_{L^\infty(\Omega)} \leq 1}} \left(\int_\Omega u \ \nabla \cdot g \ dx \right) \tag{2.47}$$

Figure 2.4: Functions with bounded and unbounded total variation



(a) $sin(\frac{1}{x})$, unbounded TV (b) $xsin(\frac{1}{x})$, unbounded TV (c) $x^2 sin(\frac{1}{x})$, bounded TV

Besides a rule for defining the total variation, we obtain a second equivalent definition for the BV space:

Definition 2.3.6 (BV-Space)**.** *Let $\Omega \subset \mathbb{R}^m$ and $u \in L^1(\Omega)$. The function $u : \Omega \to \mathbb{R}$ is of bounded total variation in Ω, if the distributional derivative of u can be expressed by a vectorial Radon measure in Ω, i.e. if*

$$\int_\Omega u \ \frac{\partial g}{\partial x_i} \ dx \;=\; -\int_\Omega g \ dD_i u, \quad \forall g \in \mathcal{C}_0^1(\Omega), \quad i = 1, ..., d \tag{2.48}$$

holds, where $Du = (D_1 u, ..., D_d u)$ is a measure with values in Ω.
The space of functions with bounded variation is defined as

$$BV(\Omega) \;:=\; \{u \in L^1(\Omega) \mid \frac{\partial u}{\partial x_i} \text{ is a Radon measure, } i = 1, ..., d\}, \tag{2.49}$$

with the norm

$$||u||_{BV(\Omega)} \;:=\; \int_\Omega |u(x)|\; dx \;+\; |u|_{TV}, \tag{2.50}$$

with the semi-norm (or total variation)

$$|u|_{TV} \;:=\; TV[u] \;=\; \int_\Omega |Du| \;=\; \sqrt{\sum_{i=1}^{d} \left(\int_\Omega d|D_i u| \right)^2}\,. \tag{2.51}$$

As indicated above, BV extends the Sobolev space $W^{1,1}(\Omega)$. Taking a function u in $W^{1,1}(\Omega)$ implies

$$\int_\Omega |Du| \;=\; \sup_{\substack{g\in\mathcal{C}_0^1(\Omega)\\ ||g||_{L^\infty(\Omega)}\le 1}} \left(\int_\Omega g\cdot\nabla u\; dx \right) \;\le \int_\Omega |\nabla u|\; dx\;. \tag{2.52}$$

Following Lusin and Uryson one can prove the converse inclusion, such that

$$\int_\Omega |Du| = \int_\Omega |\nabla u|\; dx. \tag{2.53}$$

Hence the total variation generalizes the Sobolev norm and we can conclude

$$W^{1,1}(\Omega) \subseteq BV(\Omega)\;. \tag{2.54}$$

Remembering the example of a Heaviside jump function, the space $BV(\Omega)$ is actually large enough to regard piecewise constant functions. In particular for images with sharp edges edges the BV space is a better choice than Sobolev spaces. Similarly, the described theory can be generalized to vector valued functions $u : \Omega \subset \mathbb{R}^d \to \mathbb{R}^n$.

2.3.2 Differentiability and Optimality

In the previous section we studied Banach spaces, in particular infinite dimensional function spaces, which are of special interest in image processing and imaging. The calculus of variations mainly deals with questions like existence, uniqueness or the computation of stationary points of functionals in Banach spaces. The aim of this section is to recall basic tools for derivatives and optimality in variational problems. The calculus of variations can be seen as a generalization of extreme values of functions. Thus, we first derive basic concepts of derivatives and gradients for functionals.

Definition 2.3.7 (Directional derivative)**.** *Let $J : \mathcal{U} \to \mathcal{V}$ be a functional between Banach spaces $\mathcal{U}$ and $\mathcal{V}$. The directional derivative (also called first variation) at $u \in \mathcal{U}$ in direction $v \in \mathcal{U}$ is defined as*

$$d_v J(u) \;:=\; \lim_{t\downarrow 0} \frac{J(u+tv) \;-\; J(u)}{t}\;,$$

if that limit exists.

In practice, for the computation of a directional derivative it is useful to define a function $\psi_v(t) := J(u+tv),\ t \geq 0$ for an arbitrary direction $v \in \mathcal{U}$. For an optimal u it holds $\psi_v(0) \leq \psi_v(t)$. Hence the directional derivative of J is given by $\psi_v'(t)|_{t=0}$.

Similarly to the definition of the directional derivative one can define higher order variations. For example the second variation is defined as follows

Definition 2.3.8 (Second variation)**.** *Let $J : \mathcal{U} \to \mathcal{V}$ be a functional between Banach spaces $\mathcal{U}$ and $\mathcal{V}$ and let the first variation $d_v J(u)$ exist. The second directional derivative (also called second variation) at position u in direction w is defined as*

$$d^2_{v,w} J(u) := \lim_{t \downarrow 0} \frac{d_v J(u+tw) - d_v J(u)}{t} ,$$

if that limit exists.

In the following we shortly characterize two types of differentiability, namely Gâteaux- respectively Fréchet-differentiability.

Definition 2.3.9 (Gâteaux-differentiability)**.** *Let $J : \mathcal{U} \to \mathcal{V}$ be a functional between Banach spaces $\mathcal{U}$ and $\mathcal{V}$. The set*

$$dJ(u) := \{\, d_v J(u) < \infty \mid v \in \mathcal{U} \,\} ,$$

is called Gâteaux-derivative. J is called Gâteaux-differentiable, if the set is not empty.

Furthermore, we are interested in cases where the Gâteaux-derivative is a singleton. This leads to the term of Fréchet-differentiability.

Definition 2.3.10 (Fréchet-differentiability)**.** *Let $J : \mathcal{U} \to \mathcal{V}$ be a functional with Banach spaces $\mathcal{U}$ and $\mathcal{V}$, and let $d_v J(u)$ exist for all $v \in \mathcal{U}$. If there exists a continuous linear functional $J'(u) \in \mathcal{U}$, such that*

$$J'(u)v = d_v J(u), \quad \forall v \in \mathcal{U}, \tag{2.55}$$

and

$$\frac{\|J(u+v) - J(u) - J'(u)v\|_{\mathcal{V}}}{\|v\|_{\mathcal{U}}} \longrightarrow 0 , \qquad \text{for } \|v\|_{\mathcal{U}} \to 0$$

hold, then J is called Fréchet-differentiable in u and J' is called Fréchet-derivative.

If J is Fréchet-differentiable at u, then for all directions $v \in \mathcal{U}$ there exists a Gâteaux-derivative concerning the first condition in (2.55). Analogously, a functional J is

called twice Fréchet-differentiable, if the properties in the definition are fulfilled for $d^2_{v,v}J(u)$ as well. In this case we will simply write J'' for the second Fréchet-derivative.

Besides computing potential extremal functions the question of existence and uniqueness arises. The question of existence can be answered via the fundamental theorem of optimization. First, we have to define the term of lower semi-continuity in the special case of a Banach space.

Definition 2.3.11 (Lower semi-continuity)**.** *Let $\mathcal{U}$ be a Banach space with topology τ. The functional $J : (\mathcal{U}, \tau) \to \bar{\mathbb{R}}$ is called lower semi-continuous at $u \in \mathcal{U}$ if*

$$J(u) \leq \liminf_{k\to\infty} J(u_k) \ ,$$

for all $u_k \to u$ in the topology τ.

Together with compactness this leads to the fundamental theorem of optimization, see [197, 7].

Theorem 2.3.12 (Fundamental theorem of optimization)**.** *Let $\mathcal{U}$ be a Banach space with topology τ and let $J : (\mathcal{U}, \tau) \to \bar{\mathbb{R}}$ be lower semi-continuous. Furthermore let the level set*

$$\{u \in \mathcal{U} \mid J(u) \leq C\}$$

be non-empty and compact in the topology τ for some $C \in \mathbb{R}$. Then there exists a global minimum of

$$J(u) \longrightarrow \min_{u\in\mathcal{U}} \ .$$

Proof. Let $\tilde{J} = \inf_{u\in\mathcal{U}} J(u)$. Then a subsequence $(u_k)_{k\in\mathbb{N}}$ exists with $J(u_k) \to \tilde{J}$ for $k \to \infty$. For k sufficiently large, $J(u_k) \leq C$ holds and hence $(u_k)_{k\geq k_0}$ is contained in a compact set. As a result, a subsequence $(u_{k_l})_{l\in\mathbb{N}}$ exists with $u_{k_l} \to \tilde{u}$ for $l \to \infty$ for some $\tilde{u} \in \mathcal{U}$. From the lower semi-continuity of J we obtain

$$\tilde{J} \ \leq \ J(\tilde{u}) \ \leq \ \liminf_{k\to\infty} J(u_k) \ \leq \ \tilde{J} \ ,$$

consequently $\tilde{u}$ is a global minimizer. □

In finite dimensional optimization this theorem can simply be used for existence proofs. In that case, boundedness yields compactness. However, in function spaces this is not the case due to the infinite dimension. To deduce compactness from boundedness we need weaker topologies, the so-called weak respectively weak* topologies.

Definition 2.3.13 (Weak and weak* topology)**.** *Let $\mathcal{X}$ be a Banach space and let $\mathcal{X}^*$ denote its dual space. Then the weak topology on $\mathcal{X}$ is defined as*

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

and the weak topology on X^* is defined as*

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

The weak* topology on $\mathcal{X}^*$ is weaker than the weak topology on $\mathcal{X}^*$ since $\mathcal{X} \subset \mathcal{X}^{**}$. For a reflexive Banach space ($\mathcal{X} = \mathcal{X}^{**}$), the weak and weak* topology coincide. The fundamental result for attaining compactness is the theorem of Banach-Alaoglu, which deduces compactness from boundedness at least in the weak* topology:

Theorem 2.3.14 (Theorem of Banach-Alaoglu)**.** *Let $\mathcal{X}$ be a Banach space, $\mathcal{X}^*$ its dual space and $C > 0$. Then the set*

$$\{v \in \mathcal{X}^* \mid \|v\|_{\mathcal{X}^*} \leq C\}\,, \qquad \text{for } C > 0 \ ,$$

is compact in the weak topology.*

Hence, we can prove existence of a global minimum for a given infinite dimensional optimization problem, if we are able to prove lower semi-continuity in the weak* topology. In general, proving lower semi-continuity is not an easy task. If we have existence, then we can simply compute the Fréchet-derivative to obtain a minimum.

3 Convex Analysis and Bregman Distances

In this chapter we will review basic concepts of convex analysis including subdifferentials and Legendre-Fenchel duality. Subsequently, we will introduce Bregman distances, their properties and an overview of their usage in this work.

3.1 Introduction

In this introductory section we review techniques from convex analysis used in this thesis, basically concerning subdifferential calculus and the Legendre Fenchel duality concept. Standard textbooks dealing with these topics are e.g. [154],[62], [12], [105], [100, Chapter 4], [8], [155],[23], [101] or [24].

3.1.1 Subdifferential Calculus

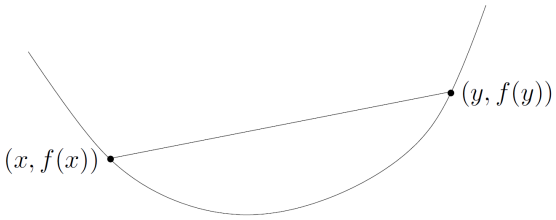
In the following we assume that X is a Banach space and, that $J : X \to [0, \infty]$ is a convex and proper penalty functional.

Definition 3.1.1 (Convexity)**.** *An operator $J : X \to [0, \infty]$ is called convex, if for $u, v \in X$ and $\lambda \in [0, 1]$ the following inequality holds:*

$$J(\lambda u + (1 - \lambda)v) \;\leq\; \lambda J(u) + (1 - \lambda)J(v) \; . \tag{3.1}$$

The operator J is called strictly convex if the latter inequality is strict for all $\lambda \in (0, 1)$ and $u \neq v$.

Note that J can take the value infinity. Hence it is useful to call a functional J *proper*, if the effective domain $dom(J) := \{u \in X \mid J(u) < \infty\} \neq \emptyset$. The effective domain of a convex functional is always a convex set.



First- and second-order conditions for convexity: The following Lemma provides a first-order condition for convexity.

Lemma 3.1.1. *Let J be differentiable over its domain. Then, J is convex if and only if its domain is convex and*

$$J(v) \geq J(u) + DJ(u)(v-u), \quad \forall\, u, v \in dom(J)\ .$$

In other words, for a convex J, a first-order approximation, i.e. a Taylor linearization, provides a global underestimate of J. Let us assume that J is twice differentiable, i.e. its Hessian or second derivative D^2u exists at each point in $dom(J)$, which is open. Then, a second-order condition for convexity is given by the following Lemma.

Lemma 3.1.2. *J is convex if and only if $dom(J)$ is convex and its Hessian is positive semidefinite:*

$$D^2J(u) \succeq 0, \quad \forall\, u \in dom(J)\ .$$

The geometric interpretation of this convexity condition is that the graph of J is required to have positive (upwards) curvature at u.

Examples: For $X = \mathbb{R}$ and $J : \mathbb{R} \to [0, \infty)$ powers of the absolute value function $J(u) := |u|^p, p \geq 1$, depicted in Fig. 3.1(a) for $p = 1$, are canonical examples of convex functions. Another example is a quadratic-over-linear, see [24, page 73] function. For $X = \mathbb{R} \times \mathbb{R}^+$ and $J : \mathbb{R} \times \mathbb{R}^+ \to [0, \infty]$ the function $J(u, v) := \frac{u^2}{v}$ is convex, see Fig. 3.1(b). This can simply be verified by

$$D^2J(u,v) = \frac{2}{v^3}\begin{pmatrix} v^2 & -uv \\ -uv & u^2 \end{pmatrix} = \frac{2}{v^3}\begin{pmatrix} v \\ -u \end{pmatrix}\begin{pmatrix} v \\ -u \end{pmatrix}^T \succeq 0\ , \tag{3.2}$$

for all $v > 0$, i.e. the Hessian of J is positive semidefinite. A variant of a quadratic-over-linear function in a functional setting will play a role in our later 4D imaging model in Chapter 7. In the case where X is the function space BV the total variation functional, we introduced in Section 2.3.1,

$$J(u) \;=\; |u|_{BV(\Omega)} \;=\; \sup_{\substack{g \in C_0^\infty(\Omega, R^d) \\ ||g||_\infty \leq 1}} \int_\Omega u\ \nabla \cdot g\ , \tag{3.3}$$

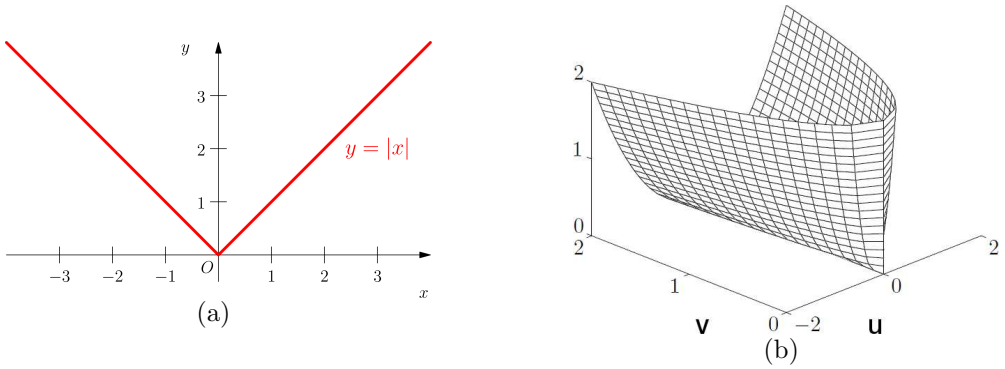


Figure 3.1: Examples of convex functions: (a) absolute value function $J(u) = |u|$ and (b) $J(u) = \frac{u^2}{v}$.

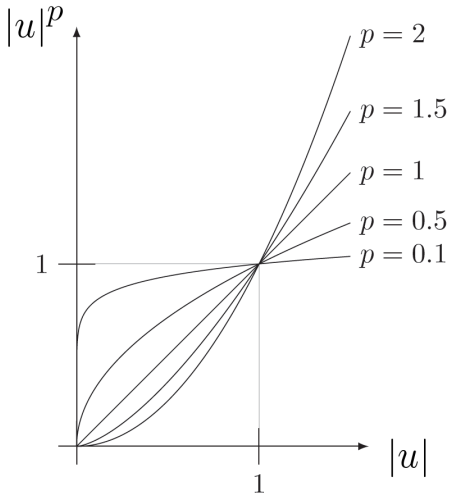


Figure 3.2: Quasiconvex functions $J(u) = |u|^p$ on the positive axis with $p \in (0, 1)$

is a prominent convex representative. In recent years sparsity models became very popular in imaging and image processing. In such cases one often has to face non-convex optimization problems. For example think of problems including $|u|^p$, with $p \in (0, 1)$. For such problems a generalized convexity concept turns out to be useful. Quasiconvexity is a generalized concept, which is defined as follows:

Definition 3.1.2 (Quasiconvexity)**.** *Let J be defined on a convex set $X \subseteq \mathbb{R}^d$. The function J is said to be quasiconvex on X if*

$$J(\lambda u + (1 - \lambda)v) \;\leq\; \max\{\, J(u)\,,\; J(v)\,\}\;,$$

for every $u, v \in X, u \neq v$, and for every $(0, 1)$. If the inequality is strict, the function is called strictly quasiconvex. .

The functions we illustrated in Figure 3.2 are strictly quasiconvex. In denoising

problems with sparsity regularization, one often considers l^p-penalized Tikhonov functionals corresponding to Figure 3.2. The existence of minimizers for such problems can be guaranteed for $p \in (0, 1)$. However, a minimizer needs not to be unique. For $p = 0$ the existence of a solution for such sparsity problems is not assured. For further details on generalized convexity we refer e.g. to [38, Chapter 2], [170] or [24, Section 3.4].

In order to obtain optimality conditions for general variational problems in imaging, see (2.12), we need a general notion of derivatives. In particular this is required for non-differentiable functionals, e.g. for the total variation, which is not differentiable in the classical sense. In the following we introduce the notion of a subdifferential (cf. [63]).

Definition 3.1.3 (Subdifferential)**.** *Let $J : X \to \bar{\mathbb{R}}$ be a convex functional and let X^* denote the dual space of X. J is called subdifferentiable at $u \in X$ if there exists an element $p \in X^*$ such that*

$$J(v) - J(u) - \langle p, v - u \rangle \;\geq\; 0 \;,\; \forall v \in X \;.$$

We call the generalized derivative p a subgradient of J at position u. The collection of all subgradients

$$\partial J(u) \;:=\; \{p \in X^* \;:\; J(v) - J(u) - \langle p, v - u \rangle \;\geq\; 0 \;,\; \forall v \in X\} \subset X^* \qquad (3.4)$$

is called subdifferential of J at u. The effective domain of the subdifferential is given by

$$D(\partial J) \;=\; \{u \in X \;:\; \partial J(u) \neq \emptyset\} \;.$$

A subgradient $p \in \partial J(u)$ can be identified with the slope of a plane (of codimension one) in $X \times \mathbb{R}$ through $(u, J(u))$, which lies under the graph of J. If J is Frechét differentiable, the subgradient of J is a singleton and coincides with its Frechét derivative, i.e. $\partial J(u) = \{\nabla J(u)\}$.

An illustrative example is the subdifferential of the absolute value function from above in Figure 3.1(a).

Example 3.1.4 (Subdifferential of the absolute value)**.** *Let $X = \mathbb{R}$ and $J : \mathbb{R} \to [0, \infty), J(u) := |u|$, be the absolute value function. Then the subdifferential of J at u is given by*

$$\partial J(u) \;=\; \begin{cases} \{1\}\,, & \text{for } u > 0 \\ [-1, 1], & \text{for } u = 0 \\ \{-1\}\,, & \text{for } u < 0 \;. \end{cases}$$

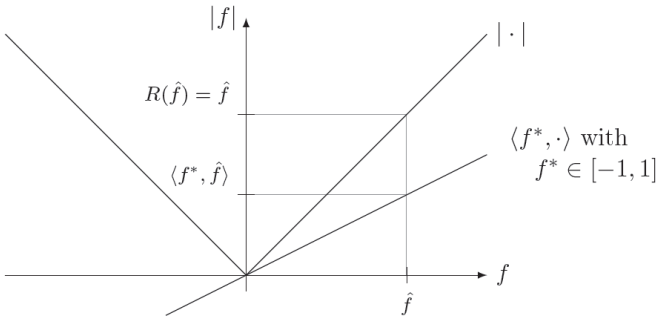


Figure 3.3: Visualization of subgradients $p \in \partial|u|$ of the absolute value function.

We plotted subgradients of the absolute value function in Figure 3.3.

In this work the total variation semi-norm will be a fundamental regularization functional. Thus, we will characterize the subdifferential of the total variation in the following, which is not an easy task. For this purpose the one-homogeneity of the total variation is a useful property. Hence, we start with characterizing the subdifferential of one-homogeneous functionals,

Lemma 3.1.3. *Let $J : X \to \bar{\mathbb{R}}$ be a convex homogeneous functional of degree one, i.e. $J(\alpha\, u) = \alpha\, J(u),\ \forall\, \alpha > 0$. Then the subdifferential of J is given by*

$$\partial J(u) = \{\, p \in X^* \mid \langle p, u \rangle = J(u),\ \langle p, v \rangle = J(v), \quad \forall v \in X \,\} \ . \tag{3.5}$$

Proof. The definition of the subgradient concerning J yields

$$\langle p, v - u \rangle \leq J(v) - J(u),$$

for all v in X. Using $v = 0$ we find

$$\langle p, u \rangle \geq J(u) \ ,$$

and using $v = 2u$ we find

$$\langle p, u \rangle \leq J(2u) - J(u) = 2J(u) - J(u) = J(u) \ ,$$

where we applied the one-homogeneity of J. Thus, we obtain $\langle p, u \rangle = J(u)$ and the assertion in (3.5) follows by one-homogeneity of J. □

In the case of total variation we see that for each subgradient p the dual norm is bounded by

$$\|p\| = \sup_{\substack{v \in BV_0(\Omega) \\ |v|_{BV(\Omega)}=1}} \langle p, v \rangle \leq \sup_{\substack{v \in BV_0(\Omega) \\ |v|_{BV(\Omega)}=1}} |v|_{BV(\Omega)} = 1 \ .$$

Hence, this yields

$$\partial |u|_{BV(\Omega)} = \left\{ p \in X^* \mid \|p\|_\infty = 1,\ \langle p, u \rangle = |u|_{BV(\Omega)} \right\} .$$

From the structure of $(BV(\Omega))^*$ we see that for each subgradient $p \in X^*$ with $\|p\| \leq 1$ there exists a vector field $g \in L^\infty(\Omega; \mathbb{R}^d)$ with $p = \nabla \cdot g$ (the opposite is not true in general). As a consequence, we can characterize the subdifferential of the total variation as follows,

Example 3.1.5 (Subdifferential of TV)**.**

$$\partial |u|_{BV(\Omega)} = \left\{ \nabla \cdot g \mid \|g\|_\infty \leq 1,\ \langle \nabla \cdot g, u \rangle = |u|_{BV(\Omega)},\ g|_{\partial\Omega} = 0 \right\} .$$

In order to obtain optimality conditions for a general convex variational problem we can use the subdifferential.

Theorem 3.1.6 (Subdifferential and Optimality)**.** *Let J be a convex functional and X a Banach space. An element $u \in X$ is a minimizer of J if and only if $0 \in \partial J(u)$.*

Proof. Let $0 \in \partial J(u)$, then the definition of a subgradient yields

$$0 \ = \ \langle 0, v - u \rangle \ \leq \ J(v) - J(u) \quad \forall v \in X \ ,$$

and hence u is a global minimizer of J. And vice versa if $0 \notin \partial J(u)$ there exists at least one $v \in X$ with

$$J(v) - J(u) \ < \ \langle 0, v - u \rangle \ = \ 0$$

so that u cannot be a minimizer of J. □

Due to convexity the first-order optimality condition is not only necessary, but also sufficient.

We presented the computation of the subdifferential of TV above. However, standard variational problems in image reconstruction usually consist of a data fidelity functional $H_f(Ku)$ and a regularization functional $J(u)$, see (2.12). Computing the subdifferential $\partial(H_f(Ku) + J(u))$ is not always trivial. With an additional regularity on H_f, e.g. if H_f is continuous or differentiable, then the following equality holds

$$\partial(H_f(u) + J(u)) \ = \ \partial H_f(Ku) + \partial J(u) \ .$$

For a proof and further details we refer to [62, Chapter 1, Proposition 5.6].

3.1.2 Legendre-Fenchel Duality

In this subsection we review concepts of Legendre-Fenchel duality. In the realm of function spaces, convexity and duality we need a dual mapping, i.e. a pair of function spaces with an associated bilinear form. In what follows we assume that X is a Banach space and we use the notation

$$\langle x^*, x\rangle_X \;:=\; \langle x^*, x\rangle_{X^*,X} \;:=\; x^*(x)$$

for the dual pairing $\langle\cdot,\cdot\rangle : X^* \times X \to \mathbb{R}$ of X^* and X. We write X^{**} for its bidual space, i.e. $X^{**} = (X^*)^*$. For $x \in X$ the natural mapping $i_X : X \to X^{**}$ is defined by

$$i_X(x)(x^*) = \langle x^*, x\rangle_X \;.$$

The general framework for duality is built around a transform that gives an operational form to the envelope representations of convex functions. This transformation called *convex conjugate* is a generalization of the Legendre transformation. It is also known as *Legendre-Fenchel transformation* named after Adrien-Marie Legendre and Werner Fenchel. For further details on Legendre-Fenchel duality we refer to chapters in books on convex analysis [154], [155, Chapter 11], [23, Chapter 3], [24, Section 3.3 and 5].

Definition 3.1.7 (Convex conjugate)**.** *Let X be a Banach space, and let X^* be the dual space to X. For a functional $J : X \to \bar{\mathbb{R}}$ the convex conjugate $J^* : X^* \to \bar{\mathbb{R}}$ is defined by*

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

or equivalently by

$$J^*(p) := -\inf_{u\in X}\{J(u) - \langle p, u\rangle\} \;.$$

Similarly the Fenchel biconjugate is given by

$$J^{**}(u) = \sup_{p\in X^*}\{\langle p, u\rangle - J^*(p)\} = (J^*)^*(u) \;.$$

For many important functionals the Fenchel biconjugate J^{**} agrees identically with J. In general, the biconjugate is a majorant of J. For a convex functional J equality holds if and only if J is lower semi continuous or equivalently J has closed level sets. For further details on Fenchel biconjugation we refer for instance to [23, Chapter 4.2].

Restricting ourselves to the special case of the Legendre transformation for functions Fig. 3.4 offers an intuitive geometric interpretation of the transformation.

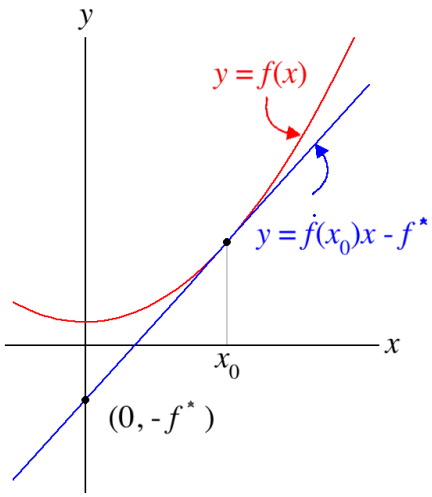


Figure 3.4: The Legendre transformation f^* of a function $f(x)$ at p_0: The function is colored in red, whereas the given dual element $p_0 = \dot{f}(x_0)$ defines the slope of the blue colored tangent at point $(x_0, f(x_0))$. The absolute value of the y-intercept is the value of the Legendre transformation $f^*(p_0)$. Indeed f^* is a maximum because every other line through f with slope p_0 will have a y-intercept above.

Theorem 3.1.8 (Fenchel´s duality theorem)**.** *Let $J_1 : X_2 \to \bar{\mathbb{R}}$ be a closed proper convex functional, let $J_2 : X_1 \to \bar{\mathbb{R}}$ be a closed proper concave functional and let $K : X_1 \to X_2$ be a bounded linear operator such that zero is in the interior, i.e. $0 \in int(dom(J_1)) - K(dom(J_2))$. Then, the following primal-dual identity*

$$\inf_{u \in X} \{ J_1(Ku) + J_2(u) \} = \sup_{p \in X^*} \{ -J_1^*(p) - J_2^*(-K^*p) \} , \tag{3.6}$$

holds.

Proof. We start with the primal problem on the left hand side,

$$\inf_u F(u) := \inf_u \{ J_1(Ku) + J_2(u) \} .$$

The latter primal problem is strongly consistent if and only if the the assumption on the interior is fulfilled. In other words we have to guarantee that $int(dom(J_1))$ and $K(dom(J_2))$ have a point in common.
By introducing a new variable $z := Ku$, we can decouple the two terms in the functional and obtain the following equivalent constraint optimization problem with two primal variables:

$$\inf_{u,z} \{ J_1(z) + J_2(u) \} \quad \text{s.t.} \quad z = Ku .$$

The corresponding Lagrangian results from introducing a Lagrange multiplier function p (dual variable) and adding a duality product to the functional. This yields a primal dual saddle point problem,

$$\inf_{u,z} \sup_{p} \{ J_1(z) + J_2(u) + \langle p, Ku - z \rangle \} .$$

We can change the order of the infimum and the supremum, and since we decoupled the system with the additional constraint we obtain an infimum for z as well as for u,

$$\sup_{p} \left\{ \inf_{z} \{ J_1(z) - \langle p, z \rangle \} + \inf_{u} \{ J_2(u) - \langle -K^*p, u \rangle \} \right\}$$
$$\sup_{p} \left\{ -\sup_{z} \{ \langle p, z \rangle - J_1(z) \} - \sup_{u} \{ \langle -K^*p, u \rangle - J_2(u) \} \right\} .$$

By using the definition of the Legendre-Fenchel transform applied to J_1^* respectively J_2^* we can get the desired dual problem

$$\sup_{p} \{ -J_1^*(p) - J_2^*(-K^*p) \} .$$

□

Remark. In Chapter 2 we introduced general variational problems for image reconstruction. Mainly, they consist of a data fidelity H_f dependent on a linear operator K and a regularization term J. Note that we can simply apply the Fenchel duality Theorem 3.1.8 with $J_1 := H_f$ and $J_2 := -J$ to obtain a corresponding dual variational formulation. In Subsection 3.3.3 we will use the key ideas of the latter proof to derive the dual Bregman algorithm. For further insights on the Legendre-Fenchel duality theorem and extensions we refer to [154, Section 31] and [155].

Table 3.1: Table of functionals and their convex conjugates

$J(u)$	$J^*(p)$
$\|u\|_2$	$\chi_{\{p \mid \|p\|_2 \leq 1\}}$
$\frac{1}{2\alpha}\|u\|_2^2$	$\frac{\alpha}{2}\|p\|_2^2$
$\|u\|_1$	$\chi_{\{p \mid \|p\|_\infty \leq 1\}}$
$\|u\|_\infty$	$\chi_{\{p \mid \|p\|_1 \leq 1\}}$
$\max(u)$	$\chi_{\{p \mid p \geq 1 \wedge \|p\|_1 = 1\}}$

3.2 Bregman distances

The Bregman distance is named after L. M. Bregman, who introduced the concept in 1967 [27]. The Bregman distance for general convex, not necessarily differentiable functionals, is defined as follows, see e.g. [45, 111].

Definition 3.2.1 (Bregman Distance)**.** *Let X be a Banach space and $J : X \to \bar{\mathbb{R}}$ be a convex functional with non-empty subdifferential ∂J. Then, the Bregman distance is defined as*

$$D_J^{\partial J(v)}(u, v) \;:=\; \{J(u) - J(v) - \langle p, u - v\rangle \,|\, p \in \partial J(v)\} \;.$$

The Bregman distance for a specific subgradient p is defined as $D_J^p : X \times X \to \mathbb{R}^+$ with

$$D_J^p(u, v) \;:=\; J(u) - J(v) - \langle p, u - v\rangle \,, \qquad p \in \partial J(v) \,,$$

where $\langle a, b\rangle \;:=\; \langle a, b\rangle_X$ denotes the dual product for $a \in X^*$ and $b \in X$ for the sake of simplicity.

For a continuously differentiable functional there is a unique element in the subdifferential and consequently a unique Bregman distance. Intuitively, the Bregman distance can be interpreted as the difference between the value of J at u and the value of the first-order Taylor expansion of J around u^l evaluated at u, see Figure 3.5. In other words, the Bregman distance can be seen as a tail of a Taylor linearization.

Example: Total Variation
The Bregman distance with respect to the total variation semi-norm $|\cdot|_{BV(\Omega)}$ is defined as follows:

$$D^p_{|\cdot|_{BV(\Omega)}}(u, v) \;:=\; |u|_{BV(\Omega)} \;-\; |v|_{BV(\Omega)} \;-\; \langle p, u - v\rangle \;,$$

where $p \in \partial\, |v|_{BV(\Omega)} \subseteq (BV(\Omega))^*$ is a subgradient in the dual space of BV.

Properties of the Bregman distance:
The Bregman distance is a distance in the sense that for $p \in \partial J(v)$

$$D_J^p(u, v) \begin{cases} = 0, & \text{if } u = v \\ \geq 0, & \text{else} \end{cases}$$

holds.

Positivity: The positivity is due to the convexity of J. If J is strictly convex, we even obtain $D_J^p(u, v) > 0$ for $u \neq v$.

Convexity: The Bregman distance is convex in the first argument, but not necessarily in the second argument.

Conjugate Duality: Let the convex functional J have a convex conjugate J^*. The Bregman distance defined with respect to J^* has an interesting relationship to D_J,

$$\sup_v D_J^p(u, v) = \sup_q D_{J^*}^u(p, q) ,$$

where $p \in \partial J(v)$ and $u \in \partial J^*(q)$.

Proof. For $p \in \partial J(v)$ and $u \in \partial J^*(q)$ we can apply the Legendre-Fenchel transform twice and obtain

$$\begin{aligned}
\sup_v D_J^p(u, v) &= \sup_v \{J(u) - J(v) - \langle p, u - v\rangle\} \\
&= J(u) - \langle p, u\rangle + \sup_v \{\langle p, v\rangle - J(v)\} \\
&= \sup_q \{\langle u, q\rangle - J^*(q)\} - \langle p, u\rangle + J^*(p) \\
&= \sup_q \{J^*(p) - J^*(q) - \langle u, p - q\rangle\} = \sup_q D_{J^*}^u(p, q) .
\end{aligned}$$

□

In Section 3.3.3 we will introduce a dual Bregman iteration. This property is one explanation why the dual Bregman iteration produces similar results compared to the primal Bregman iteration, particularly in the case of L^2 data fidelities.

In general, the Bregman distance is not a metric, since no triangular inequality nor symmetry holds for the Bregman distance. The latter one can be achieved by introducing the so-called symmetric Bregman distance.

Definition 3.2.2 (Symmetric Bregman Distance)**.** *Let X be a Banach space and $J : X \to \bar{\mathbb{R}}$ be a convex functional with non-empty subdifferential ∂J. Then, a symmetric Bregman distance is defined as $D_J^{symm} : X \times X \to \mathbb{R}^+$ with*

$$\begin{aligned}
D_J^{symm}(u_1, u_2) &:= D_J^{p_1}(u_2, u_1) + D_J^{p_2}(u_1, u_2) \\
&= \langle u_1 - u_2, p_1 - p_2\rangle_{X^*} ,
\end{aligned}$$

with $p_i \in \partial J(u_i)$ *for* $i \in \{1, 2\}$.

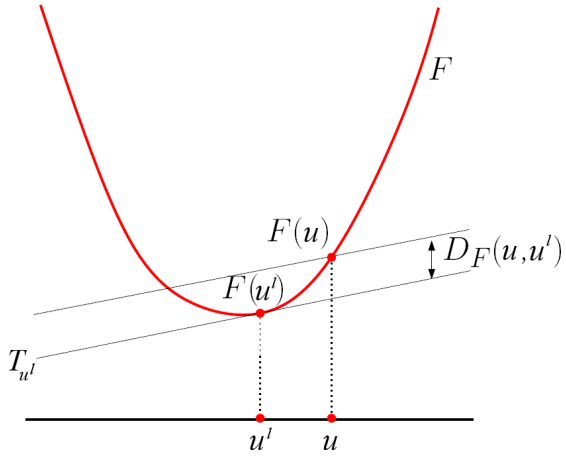


Figure 3.5: Quadratic function and Bregman distance as L^2 norm

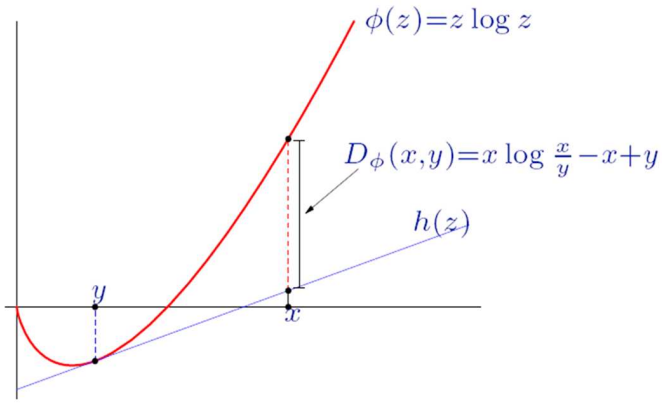


Figure 3.6: The Kullback-Leibler functional is a Bregman distance

Passing over to the symmetric Bregman distance of the convex conjugate J^*, obviously the following identity holds concerning duality

$$D_J^{\text{symm}}(u_1, u_2) \;=\; D_{J^*}^{\text{symm}}(p_1, p_2) \;.$$

Bregman distance and underlying functionals

The squared L^2 norm is a Bregman distance with the differentiable and strictly convex functional $J(u) = ||u||^2$ as the underlying functional, see Figure 3.5. Another example of a Bregman distance is the relative entropy or Kullback-Leibler divergence. The latter is generated by the strictly convex and differentiable Shannon entropy which is defined on the positive domain, see Figure 3.6. The latter distances can be used as data fidelities for Gaussian and Poisson noise modeling. Further Bregman distances and their underlying functions are presented in Table 3.7.

Function Name	$\phi(x)$	$\operatorname{dom}\phi$	$D_\phi(x;y)$
Squared norm	$\frac{1}{2}x^2$	$(-\infty,+\infty)$	$\frac{1}{2}(x-y)^2$
Shannon entropy	$x\log x-x$	$[0,+\infty)$	$x\log\frac{x}{y}-x+y$
Bit entropy	$x\log x+(1-x)\log(1-x)$	$[0,1]$	$x\log\frac{x}{y}+(1-x)\log\frac{1-x}{1-y}$
Burg entropy	$-\log x$	$(0,+\infty)$	$\frac{x}{y}-\log\frac{x}{y}-1$
Hellinger	$-\sqrt{1-x^2}$	$[-1,1]$	$(1-xy)(1-y^2)^{-1/2}-(1-x^2)^{1/2}$
ℓ_p quasi-norm	$-x^p$ $(0<p<1)$	$[0,+\infty)$	$-x^p+p\,xy^{p-1}-(p-1)\,y^p$
ℓ_p norm	$\lvert x\rvert^p$ $(1<p<\infty)$	$(-\infty,+\infty)$	$\lvert x\rvert^p-p\,x\operatorname{sgn}y\lvert y\rvert^{p-1}+(p-1)\lvert y\rvert^p$
Exponential	$\exp x$	$(-\infty,+\infty)$	$\exp x-(x-y+1)\exp y$
Inverse	$1/x$	$(0,+\infty)$	$1/x+x/y^2-2/y$

Figure 3.7: Overview of functions and their Bregman distances, see [11]

Bregman distances in this work

In this work Bregman distances will be used in several parts for different applications, playing different roles. In the following we will summarize the main applications of Bregman distances in this work.

First Application: An iterative regularization technique

In recent years Bregman distances have been used successfully for iterative regularization methods. Particularly for the case of L^2 data fidelities it has been shown that such methods can significantly improve reconstruction results. For example in the case of total variation, iterative Bregman regularization simultaneously enhances contrast in reconstructions by adding back residuals. In the following section we will concentrate on iterative Bregman regularization techniques. Particularly, a new dual Bregman iteration scheme can handle general data fidelities and has nice properties. A justification for the "strength" of Bregman regularization is given by the concept of error forgetting.

Second Application: An analytical tool

In regularization theory in Banach spaces the Bregman distance plays an important role. It has proven to be a valuable tool to study convergence and convergence rates of various regularization problems, see for example [102], [149], [152] or [36]. Among other things, the success of this technique is due to the fact that the Bregman distance automatically provides the suitable topology subordinate to a given problem. For example we automatically obtain a strong topology in L^2 or a weak* topology

in BV. A major step for error estimates and multi-scale techniques in the case of regularization with singular energies has been the introduction of (generalized) Bregman distances (cf. [27, 45, 111]) as an error measure (cf. [36]). In this work we will use Bregman distances for different analytical purposes. For example we will derive error estimates respectively convergence rates for a new dual Bregman iteration scheme via Bregman distances, see Section 3.3.3. Furthermore, we will make use of Bregman estimates to prove the existence of a solution of a new 4D TV reconstruction model in Chapter 10.

Third Application: A key to fast L^1-type algorithms
For L^1-type problems Bregman techniques turned out to be very useful to derive fast algorithms (e.g. Split Bregman resp. Douglas-Rachford splitting). The main idea of these splitting methods is to decouple functionals by substituting specific terms. Hence resulting splitting schemes only depend on fast and easy computable substeps (shrinkage, FFT, DCT). Another justification is given by the fact that Bregman iteration is equivalent to augmented Lagrangian methods, at least in the case of L^2 data fidelities. In Chapter 4 we will concentrate on splitting methods in convex optimization including Bregman algorithms.

3.3 Algorithms and Error Estimation

The goal of this section is to provide a general framework for simultaneous contrast enhancement in image reconstruction problems via Bregman iteration. Particularly we are interested in Bregman methods for non-standard data fidelities like the Kullback-Leibler distance resulting from Poisson noise modeling. In the following we will motivate and derive primal and dual Bregman methods in the sense of inverse scale space methods. In addition, we will prove error estimates for the dual Bregman method and we will explain why Bregman forgets errors. This section is general in the sense, that we will allow general convex data fidelities and general convex, usually one-homogeneous, regularization functionals. In Section 5.6 we will apply these techniques to 3D image reconstruction in the case of Poisson noise and TV regularization.

3.3.1 Introduction

In the section on Bayesian modeling (2.2) we followed maximum a-posteriori probability (MAP) estimation, i.e.

$$u_{MAP} = \arg\min_u \left(-\log p(f|u) - \log p(u) \right) \ .$$

Hence, we saw that different prior densities

$$p(u) \sim e^{-\alpha\, J(u)} \ ,$$

typically Gibbs functions, lead to different regularization terms J in a variational setting. Moreover, we saw that different noise models lead to different data fidelities. For example, Gaussian noise implies an L^2 data fidelity, whereas Poisson noise yields the Kullback-Leibler functional with a strong nonlinearity in the fidelity. Hence it makes sense to use a general variational reconstruction problem with the following definitions in the course of this section

Problem 3.3.1 (Variational Problem)**.**

$$\min_{u \in \mathcal{W}(\Omega)} \{H_f(Ku - f) \ + \ \alpha\, J(u)\} \ ,$$

where

$$K : \mathcal{U}(\Omega) \to \mathcal{V}(\Sigma)$$

denotes a linear and compact operator where $\mathcal{U}(\Omega)$ and $\mathcal{V}(\Sigma)$ are Banach spaces of functions on bounded and compact sets Ω respectively Σ.

$$H_f : \mathcal{V}(\Sigma) \to \bar{\mathbb{R}}$$

is a convex data fidelity using the operator K. In order to guarantee that the data fidelity is centered at zero, we use $Ku - f$ as the argument, i.e. $H_f(Ku - f) = 0$ if $Ku = f$. This notation is particularly useful for duality arguments. Moreover,

$$J : \mathcal{W}(\Omega) \subset \mathcal{U}(\Omega) \to \bar{\mathbb{R}}$$

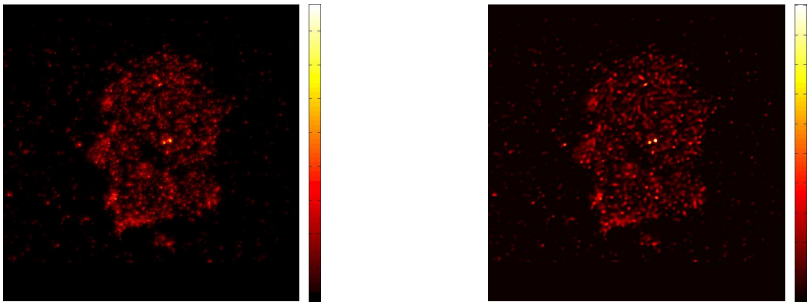
denotes a convex regularization functional. Furthermore, we call g *exact data* and f *noisy data* with a given noise estimate

$$H_f(g - f) \leq \delta \, . \tag{3.7}$$

First of all, we explain why Bregman distances respectively Bregman iteration is very useful for simultaneous contrast enhancement. To this end, let us consider regularization with total variation,

$$J(u) \ := \ |u|_{BV} \ = \ \sup_{\substack{g \in C_0^\infty(\Omega;\mathbb{R}^d) \\ ||g||_\infty \leq 1}} \int_\Omega u \, \nabla \cdot g \ .$$

As we introduced in Section 2.3.1, regularization with TV favors cartoon image reconstructions, that is it favors homogeneous regions with sharp edges. Furthermore,



Protein Syntaxin (1024x1024)

MPI Göttingen, Germany

EM-TV Reconstruction

we can expect that the length of the level sets are minimized. For example, in the figure above we illustrate a TV reconstruction of a protein in fluorescence microscopy, In the reconstruction on the right hand side, the noise has been reduced efficiently. However, we can observe a certain contrast reduction.

Systematic error of TV:
This is the well-known systematic error of total variation, cf. [124] or [134]. Although the ROF model [158], the standard reconstruction model with TV, had great success in imaging and image processing there is this deficit remaining. Compared to the original image (ground truth) we have to expect a loss of contrast in the reconstruction. This issue was studied by Meyer [124]. He verified that the application of the ROF model to the characteristic function of a ball results in a shrinked version as the minimizer u, where the shrinkage is proportional to α. In Figure 3.3.1 we illustrate the systematic error in the $1D$ case for $\alpha = 0.3$.

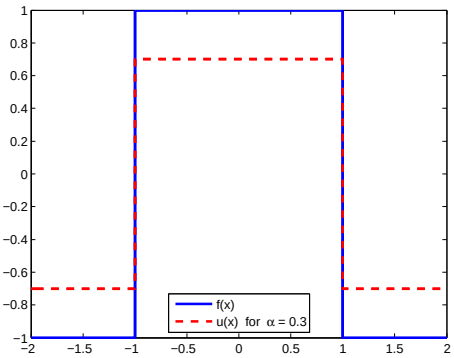


Figure 3.8: Systematic error of TV, $1D$ example for $\alpha = 0.3$

One possibility to overcome this contrast reduction is to apply iterative Bregman regularization regarding J. In the case of TV, $J := |\cdot|_{BV(\Omega)}$, we obtain the following scheme

Algorithm 3.3.2 (Primal Bregman Iteration, case H_f and TV)**.**

Let $u^0 = 0$ *and* $p^0 = 0$.
For $l = 0, 1, \ldots,$ *until stopping criterion is fulfilled:*

$$\begin{cases} u^{l+1} &= \arg\min_u \left\{ H_f(Ku - f) \;+\; \alpha\, D_J^{p^l}(u, u^l) \right\} \\ p^{l+1} &= p^l \;-\; \dfrac{1}{\alpha}\, q^{l+1} \;, \end{cases}$$

where $q^{l+1} \in \partial H_f(u^{l+1})$. The Bregman distance due to J, $D^{p^l}_{|\cdot|_{BV(\Omega)}}(u, u^l)$ is always

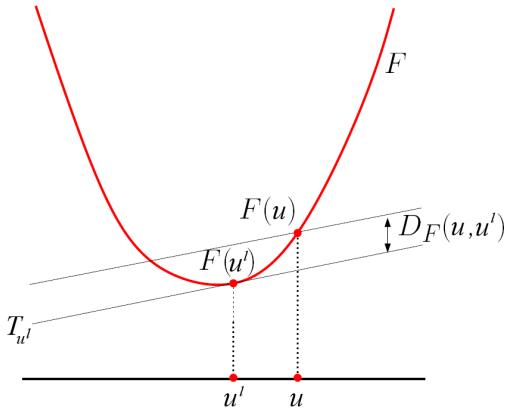


Figure 3.9: Quadratic function and Bregman distance as L^2 norm

positive, since the Bregman distance is the tail of a linearization. Furthermore, you can see in Figure 3.9 that

$$D^{p^l}_{|\cdot|_{BV(\Omega)}}(u, u^l) \;=\; 0, \quad \text{if } p^l \in \partial |u|_{BV(\Omega)} \;.$$

This is particularly the case if $u \equiv cu^l$, $\forall c \in \mathbb{R}^+$. In other words, the Bregman distance gets minimal if the edges (characterized by the subgradients) are located at the same position, whereas changes in the contrast are allowed. In the following Bregman methods we will see that contrast changes are realized by adding back residuals.

Now let us consider the following variational problem with an L^2 data fidelity as H_f and a convex regularization term $J(u)$

$$\min_u \left\{ \frac{1}{2} \|Ku - f\|^2_{L^2(\Sigma)} \;+\; \alpha\, J(u) \right\} \;,$$

Using a Bregman distance regarding the regularization functional J, the (primal) Bregman iteration for this problem reads as follows

Algorithm 3.3.3 (Primal Bregman Iteration, case L^2 and J)**.**
Let $u^0 = 0$ *and* $p^0 = 0$.
For $l = 0, 1, \ldots,$ *until stopping criterion is fulfilled:*

$$\left\{ \begin{array}{rcl} u^{l+1} & = & \arg\min\limits_u \left\{ \frac{1}{2} \|Ku - f\|^2_{L^2(\Sigma)} \; + \; \alpha \, D_J^{p^l}(u, u^l) \right\} \\ & = & \arg\min\limits_u \left\{ \frac{1}{2} \|Ku - f\|^2_{L^2(\Sigma)} \; + \; \alpha \, \left(J(u) - \langle u, p^l \rangle \right) \right\} \; , \\ p^{l+1} & = & p^l \; - \; \frac{1}{\alpha} K^*(Ku^{l+1} - f) \; , \end{array} \right.$$

where $p^l \in \partial J(u^l)$ is a subgradient in the subdifferential of J at u^l. The update of the dual variables, resp. the subgradients, directly results from the optimality condition for the variational problem with Bregman regularization

$$0 \; = \; K^*(Ku^{l+1} - f) + \alpha \, (p^{l+1} - p^l), \qquad p^{l+1} \in \partial J(u^{l+1}), \; p^l \in \partial J(u^l) \; .$$

Without affecting minimizers the scalar product in the Bregman distance in the update of u can be transferred to the L^2 data fidelity. Hence we get a shifted reference function in the data fidelity of the update of u above,

$$u^{l+1} = \; \arg\min_u \left\{ \frac{1}{2} \left\| Ku - (f + \alpha \, p^l) \right\|^2_{L^2(\Sigma)} \; + \; \alpha \, J(u) \right\} \; .$$

With the substitution

$$p^l := \frac{1}{\alpha} K^*(f^l - f) \; , \tag{3.8}$$

the primal Bregman iteration in Algorithm 3.3.3 is equivalent to the following simplified scheme without affecting minimizers.

Algorithm 3.3.4 (Primal Bregman Iteration, case L^2 and J)**.**
Let $u^0 = 0$ *and* $f^0 = f$.
For $l = 0, 1, 2, \ldots,$ *until stopping criterion is fulfilled:*

$$\left\{ \begin{array}{rcl} u^{l+1} & = & \arg\min\limits_u \left\{ \frac{1}{2} \left\| Ku - f^l \right\|^2_{L^2(\Sigma)} \; + \; \alpha \, J(u) \right\} \\ f^{l+1} & = & f^l \; + \; f - Ku^{l+1} \; , \end{array} \right.$$

with an updated reference function in the data fitting term. This reformulation is

simply possible, because

$$\begin{aligned} u^{l+1} &= \arg\min_u \left\{ \frac{1}{2} \|Ku - f\|_2^2 + \alpha D_J^{p^l}(u, u^l) \right\} \\ &\overset{(3.8)}{=} \arg\min_u \left\{ \frac{1}{2} \langle Ku, Ku \rangle - \frac{1}{2} \langle Ku, f \rangle - \frac{1}{2} \langle f, Ku \rangle \right. \\ &\qquad \left. + \frac{1}{2} \langle f, f \rangle + \alpha J(u) - \langle K^*(f^l - f), u \rangle \right\} \\ &= \arg\min_u \left\{ \frac{1}{2} \|Ku - f^l\|_2^2 + \alpha J(u) \right\} \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\alpha} K^*(f^{l+1} - f) &\overset{(3.8)}{=} p^{l+1} \overset{Alg.\ 3.3.3}{=} p^l + \frac{1}{\alpha} K^*(f - Ku^{l+1}) \\ &\overset{(3.8)}{=} \frac{1}{\alpha} K^*(f^l - f) + \frac{1}{\alpha} K^*(f - Ku^{l+1}) = \frac{1}{\alpha} K^*(f^l - Ku^{l+1}) \end{aligned}$$

hold. For further details we refer for instance to [194] or [198]. In both variants, Algorithm 3.3.3 and Algorithm 3.3.4, you can see that we add back residuals in the data fidelity. This leads to the desired simultaneous contrast enhancement in Bregman iteration. However, to transfer the shift with the scalar product in the regularization functional to the data fidelities, we made use of the L^2 structure. Thus, for general data fidelities H_f, resulting from different inverse problems and applications, the following question arises:

Is it always possible to find a shift with a residual function in the data fidelity that enhances contrast?

$$H_f(Ku - \mathbf{f^l}) \tag{3.9}$$

The answer to that question is yes and can be explained via duality. In the following sections we will introduce a new iteration based on dual Bregman regularization. We will see that this algorithm coincides with the previous (primal) Bregman iteration in the case of L^2 data fidelities.

3.3.2 Primal Bregman Iteration

In the following we present primal and dual inverse scale space strategies for solving inverse problems with general corresponding reconstruction problems. These techniques are based on iterative Bregman distance regularization for general, convex functionals and arise in the oversmoothed limit. From a dual view point of the variational model we also derive a dual inverse scale space flow, which coincides with

the primal one in the case of the Gaussian noise. In more general cases of fidelities, the dual flow appears to be easier with respect to analysis and even allows us to derive error estimates. Starting with a general, convex variational problem with data fidelity H_f and regularization functional J, we obtain the standard form for image reconstruction as written in problem 3.3.1. The corresponding iterative Bregman regularization strategy can be written as

Problem 3.3.5 (Inverse Scale Space)**.**

$$\begin{aligned} u^{l+1} &= \arg\min_{u\in\mathcal{W}(\Omega)} \left\{ H_f(Ku-f) \;+\; \alpha\, D_J^{p^l}(u,u^l) \right\} \\ &= \arg\min_{u\in\mathcal{W}(\Omega)} \left\{ H_f(Ku-f) \;+\; \alpha\,(J(u) - \langle u, p^l\rangle) \right\} , \end{aligned} \tag{3.10}$$

with $p^l \in \partial J(u^l)$.
The first-order optimality condition of this Bregman-regularized functional reads as follows,

$$\alpha(p^{l+1}-p^l) = -K^*(\partial H_f(Ku^{l+1}-f)) ,$$

with $p^l \in \partial J(u^l)$, and basically provides an update rule for p^l. In the limit $\alpha \to \infty$, the latter can be interpreted as a forward Euler discretization of the flow

$$\frac{d}{dt}p(t) = -K^*(\partial H_f(Ku(t)-f)) ,$$

with $p(0) = 0 \in \partial J(u(0))$, which has been termed *nonlinear inverse scale space method* (cf. [37], [33], [34]). The terminology inverse scale space method is due to the fact that this approach somehow behaves in an inverse way to the popular scale space methods (cf. [163], [192], [139]). In the case of inverse problems with Gaussian noise modeling, i.e. L^2 data fidelity, inverse scale space strategies have been well studied and error estimates could be obtained (cf. [37]). Unfortunately, in the general case above, the residuals $Ku(t) - f$ are enclosed by the derivative of the data fidelity. Unlike the special case of a L^2 data fidelity, this nonlinearity leads to mathematical difficulties if one wants to establish error estimates respectively convergence rates of the scale space method, since we need to invert ∂H_f enclosed by K^*. A different way to see the issues is to write the inverse scale-space method as

$$\partial H_f^*\left((K^*)^{-1}\frac{d}{dt}p(t)\right) = (f - Ku(t)) ,$$

with $p(0) = 0 \in \partial J(u(0))$. This is a strongly nonlinear equation for the dual variable p. We are able to overcome these difficulties in the following section by using an alternative dual scale space strategy.

3.3.3 Dual Bregman Iteration

In this subsection we are going to derive a dual inverse scale space method in terms of an iterative Bregman regularization of a dual reconstruction functional. Fortunately, it is possible to derive error estimates and convergence rates of the corresponding dual inverse scale space flow.

In order to derive the dual formulation of the Bregman regularization functional in (3.10), we use the zero centered data fidelity $H_f(\cdot)$ and introduce the convex conjugates

$$\begin{aligned} H_f^*(q) &= \sup_{v\in\mathcal{V}(\Sigma)} \left(\langle q, v\rangle_{\mathcal{V}(\Sigma)} - H_f(v)\right) \\ J^*(p) &= \sup_{u\in\mathcal{W}(\Omega)} \left(\langle p, u\rangle_{\mathcal{U}(\Omega)} - J(u)\right) . \end{aligned}$$

Under appropriate conditions, the Fenchel duality theorem (cf. [62]) implies the following primal-dual relation.

$$\begin{aligned} &\inf_{u\in\mathcal{W}(\Omega)} \left\{H_f(Ku-f) + \alpha\left(J(u) - \langle p^l, u\rangle\right)\right\} \\ &= \inf_{u,v} \sup_{q} \left\{H_f(Ku-f) + \alpha(J(u) - \langle p^l, u\rangle) + \langle v - Ku + f, q\rangle\right\} \\ &= \sup_{q} \left\{ \inf_{v}(H_f(v) + \langle v, q\rangle) + \alpha \inf_{u}(J(u) - \langle p^l + \frac{1}{\alpha}K^*q, u\rangle) + \langle f, q\rangle \right\} \\ &= \sup_{q} \left\{ -H_f^*(-q) + \langle f, q\rangle - \alpha J^*(\frac{1}{\alpha}K^*q + p^l) \right\} , \end{aligned}$$

with $\frac{1}{\alpha}K^*q + p^l \in \partial J(u) \subset \mathcal{W}(\Omega)^*$, where we have used convex conjugates and the identity $\inf(\cdot) = -\sup(-\cdot)$.

Defining $p := \frac{1}{\alpha}K^*q + p^l$, hence $q = \alpha(K^*)^{-1}(p - p^l)$ implies the dual formulation of the (primal) Bregman method above:

Problem 3.3.6 (Inverse Scale Space, dual formulation)**.**

$$p^{l+1} = \operatorname*{arg\,min}_{p\in\mathcal{U}(\Omega)^*} \left\{H_f^*(\alpha(K^*)^{-1}(p^l - p)) - \langle f, \alpha(K^*)^{-1}(p - p^l)\rangle + \alpha J^*(p)\right\} ,$$

with $\alpha(K^*)^{-1}(p^l - p) \in \partial H_f(Ku)$ and $p \in \partial J(u)$.

Now we are going to use the primal-dual relation above to provide a dual iterative Bregman regularization technique. Considering the standard regularized reconstruction model in (3.3.1), the described primal-dual relation above, with $p^l = 0$, yields the dual formulation of the variational problem:

Problem 3.3.7 (Variational Problem, dual formulation)**.**

$$\min_{p\in\mathcal{U}(\Omega)^*} \left\{\alpha J^*(p) - \langle f, \alpha(K^*)^{-1}p\rangle + H_f^*(-\alpha(K^*)^{-1}p)\right\} \ .$$

Note that the conjugate of J and the duality product act as a fidelity term and the conjugate of H_f as a regularization term in this formulation.

Consequently, the natural dual counterpart of the primal inverse scale space method, using the substitution $q := \alpha(K^*)^{-1}p$ resp. $q^l := \alpha(K^*)^{-1}p^l$, reads as follows

Problem 3.3.8 (Dual Inverse Scale Space)**.**

$$\begin{aligned} q^{l+1} &= \arg\min_{q\in\mathcal{V}(\Sigma)^*} \left\{ \alpha\, J^*(\frac{1}{\alpha}K^*(q)) - \langle f, q\rangle \ + \ D^{r^l}_{H_f^*}(-q, -q^l)\right\} \\ &= \arg\min_{q\in\mathcal{V}(\Sigma)^*} \left\{ \alpha\, J^*(\frac{1}{\alpha}K^*(q)) - \langle f, q\rangle \ + \ H_f^*(-q) + \langle r^l, q\rangle\right\} \end{aligned}$$

with $r^l \in \partial H_f^*(-q^l)$. The corresponding dual formulation of this variational problem, i.e. the primal (equal to the bidual) formulation, has a structure we are familiar with. The definition of the convex conjugate and the Fenchel duality theorem under appropriate conditions once more imply the following dual-primal relation:

$$\begin{aligned} &\inf_p \left\{\alpha J^*(p) - \langle f, \alpha(K^*)^{-1}p\rangle \ + \ H_f^*(-\alpha(K^*)^{-1}p) + \langle r^l, \alpha(K^*)^{-1}p\rangle\right\} \\ &= \inf_{p,q}\sup_v \left\{H_f^*(-q) + \langle r^l - f, q\rangle \ + \ \alpha J^*(p) + \langle K^*q - \alpha p, v\rangle\right\} \\ &= \sup_v \left\{ \inf_q (H_f^*(-q) + \langle r^l - f, q\rangle + \langle Kv, q\rangle) + \ \alpha \inf_p (J^*(p) - \langle p, v\rangle)\right\} \\ &= \sup_v \left\{ -\sup_q (H_f^*(-q) + \langle Kv + r^l - f, q\rangle) \ - \ \alpha J(v)\right\} \\ &= -\inf_v \left\{H_f(Kv + r^l - f) \ + \ \alpha J(v)\right\} \end{aligned}$$

We obtain the simple primal (bidual) iterative regularization technique, equivalent to the dual formulation in Problem 3.3.8 above:

Problem 3.3.9 (Dual Inverse Scale Space, primal formulation)**.**

$$u^{l+1} = \arg\min_{u\in\mathcal{W}(\Omega)} \left\{H_f(Ku + r^l - f) + \alpha J(u)\right\}, \tag{3.11}$$

with $r^l \in \partial H_f^*(-\alpha(K^*)^{-1}p^l)$.
Since both H_f and J are proper, lower semi-continuous and convex, and since H_f is locally bounded, we have

$$\partial\left(H_f(Ku) + \alpha J(u)\right) \;=\; \partial H_f(Ku) + \alpha\,\partial J(u)$$

for all $u \in \mathcal{W}(\Omega)$, cf. [62]. Hence, the optimality condition of (3.11) is given by

$$0 \;\in\; K^*(\partial H_f(Ku^{l+1} - f + \partial H_f^*(-\alpha(K^*)^{-1}p^l))) + \alpha p^{l+1}\ ,$$

and with the definition of convex conjugates this yields

$$(K^*)^{-1}(-\alpha p^{l+1}) \;\in\; \partial H_f(Ku^{l+1} - f + \partial H_f^*(-\alpha(K^*)^{-1}p^l))$$

and

$$\partial H_f^*(-\alpha(K^*)^{-1}p^{l+1}) \;=\; \partial H_f^*(-\alpha(K^*)^{-1}p^l) + Ku^{l+1} - f\ .$$

Consequently, the first order optimality condition of this variational problem provides an update of the residual function r^l,

$$r^{l+1} \;=\; r^l + Ku^{l+1} - f \tag{3.12}$$

for $r^l \in \partial H_f^*(-\alpha(K^*)^{-1}p^l)$ and $r^{l+1} \in \partial H_f^*(-\alpha(K^*)^{-1}p^{l+1})$. This recursion formula yields an interesting decomposition of f involving "noise" at levels l and $l+1$ and signal at level $l+1$.

Well-Definedness of the Iterates

In the following we show that the iterative dual-Bregman procedure is well-defined, i.e. that (3.12) has a minimizer u^{l+1} and that we may find a suitable subgradient r^{l+1}.

Proposition 3.3.10. *Assume H_f to be a strictly convex fidelity with operator K having a trivial null space and J to be a convex functional. Let $u^0 := 0$, $p^0 := 0 \in \partial J(u^0)$, $r^0 = 0 \in \partial H_f^*(-\alpha(K^*)^{-1}p^0)$ and $\alpha > 0$. Then, the minimizers u^{l+1} in (3.11) are well-defined.*

Proof. As described above, rewriting the optimality condition of (3.11) yields the update (3.12) of the residuals. Since $r^0 = 0$, the $l-th$ residual can be expressed explicitly by

$$r^l = -\sum_{i=1}^{l}(f - Ku^i) \in \partial H_f^*(-\alpha(K^*)^{-1}p^l),$$

consequently (3.12) changes to

$$u^{l+1} \;=\; \underset{u\in\mathcal{W}(\Omega)}{\arg\min} \left\{ H_f(Ku + \sum_{i=1}^{l}(Ku^i) - (l+1)f) + \alpha J(u) \right\} \;. \tag{3.13}$$

Hence, the existence of minimizers can be traced back to existence of minimizers for the original reconstruction problem, just with modified given data, which can be treated as usual. Moreover, as K has only a trivial null space, the strict convexity of H_f and the convexity of J imply the strict convexity of the functional (3.11), and therefore the minimizers u^{l+1} are unique. □

Dual Inverse Scale Space Flow

To derive a dual nonlinear inverse scale space flow we have to take a look at the update formula (3.12) due to the optimality condition of (3.11). In the limit $\alpha \downarrow 0$, this can be interpreted as a forward Euler discretization of the flow

$$\frac{d}{dt}r(t) \;=\; Ku(t) - f, \qquad r(0) = 0 \;, \tag{3.14}$$

with $r(t) \in \partial H_f^*(-\alpha(K^*)^{-1}p(t))$, which is termed *dual nonlinear inverse scale space method* (in analogy to previous works [34], [33]). By defining the integrated residual in that way, we obtain

$$p(t)\,\alpha \;=\; K^*(q(t)) \;. \tag{3.15}$$

Error Estimates

In order to derive error estimates in the iterative Bregman distance setting we need to introduce the so-called source condition

$$\exists\, \tilde{p} \in \partial J(\tilde{u}),\, \exists\, \tilde{q} \in \mathcal{V}(\Sigma)^* : \quad \tilde{p}\,\alpha \;=\; K^*\tilde{q}\,. \tag{SC}$$

The nowadays standard source condition (SC) will in some sense ensure that a solution $\tilde{u}$ contains features that are enhanced by the regularization term J. More precisely, for an arbitrary, but fixed $\alpha > 0$, the set of $\tilde{p}\,\alpha$ satisfying the source condition is equivalent to the set of minimizers $\tilde{p}\,\alpha$ of the corresponding variational problem, see [36, Prop. 1]. In other words, the existence of the function $\tilde{q}$ can be interpreted as the existence of a Lagrange multiplier of a constraint optimization problem, which is an additional regularity condition on the solution. In the following we will see that a source condition allows to derive a quantitative estimate between an exact solution and the solution of a regularized problem.

Notice the resemblance between the time dependent subgradient $p(t)$ in (3.15) and $\tilde{p}$ in the source condition.

The techniques we will use in the followin are related to entropy methods for PDEs, in particular they are related to the Bakry-Emery method, see e.g. [40, 41, 141, 183]. Now we consider the Bregman distance for the convex conjugate of H_f, $D_{H_f^*}^{\partial H_f^*(-q(t))}(-\tilde{q}, -q(t))$, which is finite due to the source condition. Then

$$\begin{aligned}
&\frac{d}{dt}\left(D_{H_f^*}^{\partial H_f^*(-q(t))}(-\tilde{q}, -q(t))\right) \\
=\ &\frac{d}{dt}\left(H_f^*(-\tilde{q}) - H_f^*(-q(t)) - \partial H_f^*(-q(t))(-\tilde{q} + q(t))\right) \\
=\ &\langle r_t, \tilde{q} - q(t)\rangle \overset{(3.14)}{=} \langle f - Ku(t), q(t) - \tilde{q}\rangle \\
=\ &\langle f - g, q(t) - \tilde{q}\rangle - \langle Ku(t) - g, q(t) - \tilde{q}\rangle \\
=\ &\langle f - g, q(t) - \tilde{q}\rangle - \alpha\langle u(t) - \tilde{u}, p(t) - \tilde{p}\rangle \\
=\ &\langle f - g, q(t) - \tilde{q}\rangle - \alpha D_J^{symm}(u(t), \tilde{u}) \\
\leq\ &\langle f - g, q(t) - \tilde{q}\rangle - \alpha D_J^{p(t)}(\tilde{u}, u(t)) \\
=:\ &-I(t)\ .
\end{aligned} \tag{3.16}$$

In the following we want to analyse the monotone behavior of $I(t)$. For that purpose, we deduce a relation between q_t and the second derivative of the data fidelity H_f from the dual inverse scale space flow in (3.14) first, which is

$$\begin{aligned}
\frac{d}{dt}(\partial H_f^*(-q(t))) &= \partial_q^2(H_f^*(-q(t)))\,(-q_t) \\
&= Ku(t) - f\ .
\end{aligned}$$

By using an equivalent definition of the convex conjugate in the differentiable case ($\partial H_f^* = (\partial H_f)^{-1}$) and by using the derivative of inverse functions this yields

$$\begin{aligned}
q_t(t) &= \partial_t(\alpha(K^*)^{-1}p(t)) \\
&= [\partial_q^2(H_f^*(-q(t)))]^{-1}\,(f - Ku(t)) \\
&= [\partial_u[(\partial_u H_f(\partial H_f^*(-q(t))))^{-1}]]^{-1}\,(f - Ku(t)) \\
&= \partial_u^2 H_f(r(t))\,(f - Ku(t)), \quad r(t) \in \partial H_f^*(-q(t))\ .
\end{aligned}$$

Consequently, the temporal properties of the estimate $I(t)$ read as follows:

$$\begin{aligned}
&\frac{d}{dt}\left(I(t)\right)\\
=& -\langle q_t, f-g\rangle + \alpha\frac{d}{dt}\left(D_J^{p(t)}(\tilde{u},u(t))\right)\\
=& -\langle q_t, f-g\rangle - \alpha\langle \tilde{u}-u(t), p_t\rangle\\
=& -\langle q_t, f-g\rangle - \langle g-Ku(t), \partial_t(\alpha(K^*)^{-1}p(t))\rangle\\
=& -\langle f-g, q_t\rangle - \langle g-Ku(t), H_f''(r(t))\,(f-Ku(t))\rangle\\
=& -\langle f-g, H_f''(r(t))\,(f-Ku(t))\rangle - \langle g-Ku(t), H_f''(r(t))\,(f-Ku(t))\rangle\\
=& -\langle f-Ku(t), H_f''(r(t))\,(f-Ku(t))\rangle\\
\leq& \,0
\end{aligned} \tag{3.17}$$

with $r(t)\in\partial H_f^*(-q(t))$. The latter inequality means that the Hessian H_f'' is positive semidefinite. This in turn is a second order criterion for the convexity of H_f, see Lemma 3.1.2.

Hence, after integrating inequality (3.17) from 0 to t we get a decrease of I in time and obtain

$$I(t) \;\leq\; I(s) \quad \forall t\geq s \qquad \text{hence} \qquad t\cdot I(t) \;\leq\; \int_0^t I(s)ds\;. \tag{3.18}$$

Now, integrating (3.16) from 0 to t yields

$$D_{H_f^*}^{r(t)}(-\tilde{q},-q(t)) - D_{H_f^*}^{r(0)}(-\tilde{q},-q(0)) \;\leq\; -\int_0^t I(s)ds\;,$$

which implies

$$t\cdot I(t) \overset{(3.18)}{\leq} \int_0^t I(s)ds \;\leq\; D_{H_f^*}^{r(0)}(-\tilde{q},-q(0)) - D_{H_f^*}^{r(t)}(-\tilde{q},-q(t)). \tag{3.19}$$

In the case of noise-free data, i.e. $\delta=0$, $I(t)$ reduces to the time dependent Bregman distance we want to estimate. Hence, we can conclude

$$t\,\alpha\cdot D_J^{p(t)}(\tilde{u},u(t)) \;\leq\; D_{H_f^*}^{r(0)}(-\tilde{q},-q(0)) \underbrace{-D_{H_f^*}^{r(t)}(-\tilde{q},-q(t))}_{\leq 0}$$

and thus:

Theorem 3.3.11 (Exact data)**.** *Let $\tilde{u}\in\mathcal{U}(\Omega)$ satisfy $K\tilde{u}=g=f$, and (SC). Moreover, with $q(t):=(K^*)^{-1}p(t)$, let u be a solution of the dual inverse scale space flow*

$$\partial_t\, r(t) \;=\; Ku(t)-g, \qquad r(t)\in\partial H_f^*(-q(t))\;.$$

Then the convergence rate $D_J^{p(t)}(\tilde{u}, u(t)) = \mathcal{O}(\frac{1}{t})$ *holds, more precisely*

$$D_J^{p(t)}(\tilde{u}, u(t)) \;\leq\; \frac{D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0))}{\alpha\, t} \;.$$

In the case of noisy data some further effort is necessary, since the temporal derivative of the Bregman distance with respect to H_f^* in (3.16) is not bounded by the negative Bregman distance. Then, (3.19) reads as follows:

$$D_{H_f^*}^{r(t)}(-\tilde{q}, q(t)) - \langle t(f-g), q(t) - \tilde{q}\rangle + t\,\alpha\, D_J^{p(t)}(\tilde{u}, u(t)) \leq D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0)) \;. \quad (3.20)$$

In order to find a lower bound for the first two terms on the left hand side, we provide the following lemma.

Lemma 3.3.1. *Let* $F : X \to \mathbb{R} \cup \{\infty\}$ *be a convex functional, twice continuously Frechét-differentiable,* $u, v \in X$, $p, q \in X^*$ *and* $t \in \mathbb{R}^+$. *Then the duality product can be estimated by a sum of Bregman distances,*

$$\langle t(u-v), p-q\rangle_X \;\leq\; c\, t^2\, D_F^{\partial F(u)}(v, u) + D_{F^*}^{\partial F^*(q)}(p, q),$$

with a constant,

$$c := (\inf_{w\in[v,u]} \|F''(w)\|)^{-1}\, (\inf_{\xi\in[q,p]} \left\|F^{*\prime\prime}(\xi)\right\|)^{-1} \;,$$

depending on the norm of the second derivative of F *and its convex conjugate.*

Proof. One observes from a Taylor expansion of F^* in p around q with a residual term,

$$F^*(p) \;=\; F^*(q) + \langle (F^*)'(q), p-q\rangle + \frac{1}{2}\langle p-q, (F^*)''(\xi)(p-q)\rangle$$

with $\xi \in [q, p]$, that a lower bound for the corresponding Bregman distance is given by

$$\begin{aligned} D_{F^*}^{(F^*)'(q)}(p, q) \;&=\; F^*(p) - F^*(q) - \langle (F^*)'(q), p-q\rangle \\ &=\; \frac{1}{2}\langle p-q, (F^*)''(\xi)(p-q)\rangle \\ &\geq\; \frac{\epsilon}{2}\,\|p-q\|^2, \end{aligned}$$

with $\epsilon := \inf_{\xi\in[q,p]} \|(F^*)''(\xi)\|$.
In analogy a Taylor expansion of F in v around u,

$$F(v) \;=\; F(u) + \langle F'(u), v-u\rangle + \frac{1}{2}\langle v-u, F''(w)(v-u)\rangle,$$

with $w \in [u - v]$, yields

$$\begin{aligned}\frac{1}{\tilde{c}\,\epsilon} D_F^{F'(u)}(v,u) &= \frac{1}{\tilde{c}\,\epsilon}(F(v) - F(u) - \langle F'(u), v-u\rangle) \\ &= \frac{1}{2\,\tilde{c}\,\epsilon}\langle v-u, F''(w)(v-u)\rangle \\ &\geq \frac{1}{2\,\epsilon}\|u-v\|^2 ,\end{aligned}$$

with $\tilde{c} := \inf_{w\in[u,v]} \|F''(w)\|$.
Using Young's inequality we obtain

$$\begin{aligned}\langle t(u-v), p-q\rangle &\leq \frac{t^2}{2\epsilon}\|u-v\|^2 + \frac{\epsilon}{2}\|p-q\|^2 \\ &\leq c\,t^2\, D_F^{\partial F(u)}(v,u) + D_{F^*}^{\partial F^*(q)}(p,q)\end{aligned}$$

With constant $c := \frac{1}{\tilde{c}\epsilon}$ we get the desired estimation. □

Now, applying Lemma 3.3.1 to functional H_f yields the estimate

$$\langle t(f-g), q(t) - \tilde{q}\rangle \;\leq\; D_{H_f^*}^{r(t)}(-\tilde{q}, -q(t)) + c\,t^2 D_{H_f}^{\partial H_f(f)}(g,f) \;,$$

such that the Bregman distance corresponding to H_f^* is annihilated in (3.20) and that we can conclude

$$-c\,t^2 D_{H_f}^{\partial H_f(f)}(g,f) + t\,\alpha\, D_J^{p(t)}(\tilde{u}, u(t)) \;\leq\; D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0)) \;.$$

Finally, since $H_f(f) = 0$, $\partial H_f(f) = 0$ and with the upper bound of the noise (3.7) we have

$$D_{H_f}^{\partial H_f(f)}(g,f) \;=\; H_f(g) \;\leq\; \delta \;,$$

provides a general error estimate for the dual inverse scale space method:

Theorem 3.3.12 (Noisy data)**.** *Let $\tilde{u} \in \mathcal{U}(\Omega)$ satisfy $K\tilde{u} = g$ and (SC), and let f be noisy data satisfying (3.7). Moreover, with $q(t) := (K^*)^{-1}p(t)$, let u be a solution of the dual inverse scale space flow*

$$\partial_t\, r(t) \;=\; Ku(t) - f, \qquad r(t) \in \partial H_f^*(-q(t)) \;.$$

Then the error estimate

$$D_J^{p(t)}(\tilde{u}, u(t)) \;\leq\; \frac{D_{H_f^*}^{r(0)}(-\tilde{q}, -q(0))}{\alpha\, t} \;+\; c\,\delta\, t$$

holds. In particular, for the choice $t_*(\delta) := \mathcal{O}(\frac{1}{\delta})$ we obtain the convergence rate $D_J^{p(t_*)}(\tilde{u}, u(t_*)) = \mathcal{O}(\delta)$.

Remark. In the case of Poisson noise modeling, i.e. if H_f is the Kullback-Leibler data fidelity, c_{KL} reads as follows:

$$\begin{aligned} c_{KL} &= (\inf_{w\in[g,f]} \left\|\partial_w^2 H_f(w)\right\|)^{-1} (\inf_{\xi\in[\tilde{q},q(t)]} \left\|\partial_\xi^2 H_f^*(\xi)\right\|)^{-1} \\ &= \frac{\max\{\sup g, \sup f\}^2}{\inf f} \, \frac{\max\{\sup(1-\tilde{q}), \sup(1-q(t))\}^2}{\inf f} \ , \end{aligned}$$

since

$$\partial_w^2 H_f(w) = \frac{f}{w^2}, \quad H_f^*(\xi) = \int_\Sigma -f\log(1-\xi) \quad \text{and} \quad \partial_\xi^2 H_f^*(\xi) = \frac{f}{(1-\xi)^2} \ .$$

3.3.4 Bregman and Error Forgetting

Throughout this work we will see that besides fast algorithms and analytical investigations, Bregman distances are very useful for L^1, TV and related minimization problems. Recently, Osher and Yin gave an explanation why Bregman works so well. The reason is *error cancellation* respectively *error forgetting*.

Let us consider the following variational problem with an L^2 data fidelity as H_f and a convex regularization term $J(u)$

$$\min_u \left\{ \frac{1}{2}\left\|Ku-f\right\|^2_{L^2(\Sigma)} \ + \ \alpha\, J(u) \right\} \ .$$

Using the Bregman distance from above 3.2.1 regarding the regularization functional J, the (primal) Bregman iteration for this problem reads as follows

Algorithm 3.3.13 (Primal Bregman Iteration, case L^2 and J)**.**
Let $u^0 = 0$ and $p^0 = 0$.
For $l = 0, 1, \ldots,$ until stopping criterion is fulfilled:

$$\begin{aligned} u^{l+1} &= \arg\min_{u\in\mathcal{W}(\Omega)} \left\{ \frac{1}{2}\left\|Ku-f\right\|^2_{L^2(\Sigma)} \ + \ \alpha\, D_J^{p^l}(u,u^l) \right\} \\ p^{l+1} &= p^l \ - \ \frac{1}{\alpha} K^*(Ku^{l+1}-f) \ , \end{aligned}$$

where $p^l \in \partial J(u^l)$ is a subgradient in the subdifferential of J at u^l. As described in Section 3.3.1 this Bregman iteration scheme is equivalent to

Algorithm 3.3.14 (Primal Bregman Iteration, case L^2 and J)**.**
Let $u^0 = 0$ and $f^1 = f$.
For $l = 1, 2, \ldots,$ until stopping criterion is fulfilled:

$$\begin{aligned} u^l &= \arg\min_{u\in\mathcal{W}(\Omega)} \left\{ \frac{1}{2}\left\|Ku-f^l\right\|^2_{L^2(\Sigma)} \ + \ \alpha\, J(u) \right\} \\ f^{l+1} &= f^l \ + \ f - Ku^l \ . \end{aligned} \tag{3.21}$$

In comparison to Section 3.3.1 we shifted the index of the u update here, to simplify the notation in the following part. Now we pass over to error forgetting in Bregman iteration. Let us assume we make an error ϵ^l while computing u^{l+1} in the first step of Algorithm 3.3.14, i.e.

$$\tilde{u}^l \;=\; u^l - \epsilon^l \;. \tag{3.22}$$

Thus we obtain the following disturbed update for the second step regarding f

$$\tilde{f}^{l+1} \;=\; f^l \;+\; f - K\tilde{u}^l \;,$$

and hence using the linearity of K

$$\tilde{f}^{l+1} = f^{l+1} + K\epsilon^l \;.$$

Now we have a look at the next update for u in the Bregman iteration process. Computing a new solution $\tilde{u}^{l+1}$ via the new disturbed iterate $\tilde{f}^{l+1}$ yields

$$\begin{aligned}\tilde{u}^{l+1} &\overset{(3.22)}{=} u^{l+1} \;-\; \epsilon^{l+1} \\ &\overset{(3.21)}{=} \operatorname*{arg\,min}_{u\in\mathcal{W}(\Omega)} \left\{ \frac{1}{2}\left\| Ku - \left(f^{l+1} + K\epsilon^l\right)\right\|^2_{L^2(\Sigma)} \;+\; \alpha\, J(u) \right\} \;-\; \epsilon^{l+1} \;,\end{aligned}$$

where ϵ^{l+1} denotes the new error in this minimization process. Instead of minimizing over u we minimize over $u - \epsilon^l$. This implies

$$\tilde{u}^{l+1} \;+\; \epsilon^{l+1} \;=\; \epsilon^l \;+\; \operatorname*{arg\,min}_{u\in\mathcal{W}(\Omega)} \left\{ \frac{1}{2}\left\| Ku - f^{l+1}\right\|^2_{L^2(\Sigma)} \;+\; \alpha\, J(u+\epsilon^l) \right\} \;,$$

because u and ϵ^l add in the computation of J. Since J is homogeneous of degree one we obtain

$$\begin{aligned}\tilde{u}^{l+1} \;+\; \epsilon^{l+1} \;&=\; \epsilon^l \;+\; \operatorname*{arg\,min}_{u\in\mathcal{W}(\Omega)} \left\{ \frac{1}{2}\left\| Ku - f^{l+1}\right\|^2_{L^2(\Sigma)} \;+\; \alpha\, J(u) \;+\; \alpha\, L(\epsilon^l) \right\} \\ &=\; \epsilon^l \;+\; u^{l+1} \;,\end{aligned}$$

where $L(\epsilon^l)$ is independent of u. By induction we can deduce for all l, large enough, the interesting result

$$\tilde{u}^l \;=\; u^l \;+\; \epsilon^{l-1} \;-\; \epsilon^l \;. \tag{3.23}$$

To conclude, adding back the residual in Bregman iteration does not accumulate errors. Instead the error cancels step by step. As you can see in (3.23), the magnitude of the error in l is only dependent on the magnitude of the difference of the current and previous error. In other words, Bregman iteration forgets errors by adding back residuals.

Error forgetting in Dual Bregman with general H_f:
Interestingly, error forgetting also happens in the dual Bregman iteration for general data fidelities H_f, including various nonstandard noise models,

$$\min_u \{ H_f(Ku - f) + \alpha J(u) \} .$$

In the case of an L^2 data fidelity, the primal and the dual Bregman iteration coincide. Hence, error forgetting in the dual Bregman iteration scheme happens with the same arguments as above. In the general case of dual Bregman iteration, error forgetting happens since the residuals add up in the argument of the data fidelity term, see 3.13 in the proof of the well-definedness. This is an indication that Bregman iterations can be very useful for various imaging and image processing problems with convex and 1-homogeneous regularization terms.

4 Splitting Methods

4.1 Saddle Point Problems

To describe primal-dual splitting methods for convex optimization problems we first of all introduce general saddle point problems. Various applications in literature lead to saddle point problems, e.g. computational fluid dynamics [82, 175, 190], economics [5], image registration [91, 126, 127] or parameter identification problems [35, 88, 89]. For an extensive list of applications leading to saddle point problems, we refer to the work [20].

Saddle point problems usually result from Lagrangian approaches, where Lagrange multipliers (dual variables) are used to build *unconstrained* problems from *constrained* optimization problems. Let $L : \mathcal{U} \times \mathcal{V} \to \mathbb{R} \cup \{\infty\}$ be a convex-concave functional, where $\mathcal{U}$ and $\mathcal{V}$ are Banach spaces.

Problem 4.1.1 (Primal-Dual Saddle Point Problem)**.**

$$\inf_{u \in \mathcal{U}} \sup_{p \in \mathcal{V}} L(u;p) \ , \tag{4.1}$$

where we denote u as the primal and p as the dual variable.

We are interested in finding a saddle point pair $(u^*;p^*)$ which means

$$L(u;p^*) \ \leq \ L(u^*;p^*) \ \leq \ L(u^*;p) \ , \quad \forall\, (u,p) \in \mathcal{U} \times \mathcal{V} \ . \tag{4.2}$$

By use of the dual optimality condition in (4.1) we obtain a purely primal convex problem

$$\inf_{u \in \mathcal{U}} F(u) \ , \quad \text{where } F(u) := \sup_{p \in \mathcal{V}} L(u;p) \ . \tag{4.3}$$

In an analogous way using the primal optimality condition in (4.1) we obtain a purely dual problem

$$\sup_{p\in\mathcal{V}} G(p) \ , \quad \text{where } G(p) := \inf_{u\in\mathcal{U}} L(u;p) \ . \tag{4.4}$$

Following [154, 28.3], the pair $(u^*;p^*)$ is a saddle point of the Lagrangian in (4.1), if and only if u^* is an optimal solution of the primal problem (4.3) and p^* is an optimal solution of the dual problem (4.4). In other words, the existence of a solution of the primal problem or the existence of a solution of the dual problem on its own does not imply the existence of a saddle point. The existence of a saddle point (u^*,p^*) of L implies

$$L(u^*;p^*) \ = \ \max_p \min_u \ L(u;p) \ = \ \min_u \max_p \ L(u;p) \ .$$

In Section 4.3 we will introduce augmented Lagrangian methods. A saddle point of the augmented Lagrangian is a saddle point of the Lagrangian and the converse also holds, see [83, 3.2]. For further details on saddle point problems in convex optimization we refer for example to [154, 83, 62].

4.2 Uzawa Methods

In the previous section we mentioned several applications leading to saddle point problems. In Chapter 10 we will consider a 4D image reconstruction problem, whose dynamic is modeled using a transport equation, and which leads to a saddle point formulation. Classical examples that can be formulated as saddle point problems and which are often numerically treated with Uzawa methods are the well-known Stokes problems. Hence, we will motivate and introduce general Uzawa methods via the stationary Stokes problem. For $\Omega \subset \mathbb{R}^d$ the latter reads as follows:

Problem 4.2.1 (Stationary Stokes problem)**.** *Find a solution* $(u,p) \in \left(\mathcal{C}^2(\Omega)^d \cap \mathcal{C}^1(\bar{\Omega})^d\right) \times \mathcal{C}^1(\Omega)$ *with*

$$\begin{aligned} -\Delta u + \nabla p &= f \qquad \text{in } \Omega \\ \nabla \cdot u &= 0 \qquad \text{in } \Omega \\ u &= 0 \qquad \text{in } \partial\Omega \ . \end{aligned}$$

By multiplying with testfunctions and by integrating over Ω, a weak variational formulation of the system can be written in the following way:

Find a function $u \in X := (H_0^1(\Omega))^d, p \in Y := \left\{q \in L_2(\Omega) : \int_\Omega q\, dx = 0\right\}$ with

$$\begin{aligned} \int_\Omega \nabla u \nabla v \quad - \int_\Omega \nabla \cdot v \cdot p &= \int_\Omega f \cdot v\, dx \qquad \forall v \in X \\ - \int_\Omega \nabla \cdot u \cdot q \qquad\qquad &= \qquad 0 \qquad \forall q \in Y\ . \end{aligned}$$

The first equation results from applying partial integration with zero boundary conditions. By introducing continuous bilinear forms $a(u,v) := \int_\Omega \nabla u \nabla v$ respectively $b(v,p) := -\int_\Omega \nabla \cdot v \cdot p$, and a continuous linear form on the right hand side, we obtain

$$\begin{aligned} a(u,v) \qquad\quad + b(v,p) &= \langle f, v\rangle \qquad \forall v \in X \\ b(u,q) \qquad\qquad\quad &= 0 \qquad \forall q \in Y\ . \end{aligned}$$

If the considered function space has a finite basis, a discretization yields the following system of linear equations

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}\ , \tag{4.5}$$

where we assume that A is a symmetric, positive definite $(n \times n)$ matrix, B is $(m \times m)$ matrix with full rank $m \leq n$, and B^T denotes the transposed matrix of B. Linear systems of the form (4.5) can be interpreted as Karush-Kuhn-Tucker conditions for linearly constrained quadratic programming problems or saddle point problems. Typically such systems result from finite element approximations of elliptic problems, elasticity problems or the Stokes equations as above. Solving the system (4.5) is equivalent to minimizing the following constrained optimization problem

$$\begin{aligned} &\min_u \ \frac{1}{2} u^T A u - f^T u \\ &\text{s.t. } Bu = g\ . \end{aligned}$$

This formulation reveals the correspondence to Lagrangian methods, whereas p can be interpreted as a Lagrange multiplier.

In the following we deduce the classical Uzawa algorithm via the Schur complement of A. We assume the coefficient matrix in (4.5) to be nonsingular and the (negative) Schur complement $C := BA^{-1}B^T$ of A to be symmetric and positive definite. The Schur complement arises as the result of performing a block Gaussian elimination by multiplying (4.5) from the right with the lower triangular block matrix

$$L = \begin{pmatrix} I & 0 \\ BA^{-1} & I \end{pmatrix}\ ,$$

that is we obtain

$$\begin{pmatrix} A & B^T \\ 0 & C \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} I & 0 \\ -BA^{-1} & I \end{pmatrix} \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix}$$
$$= \begin{pmatrix} I & 0 \\ -BA^{-1} & I \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix} \qquad = \begin{pmatrix} f \\ g - BA^{-1}f \end{pmatrix} .$$

Hence the reduced optimality system reads as follows

$$Au = f - B^T p \tag{4.6}$$
$$Cp = BA^{-1}f - g , \tag{4.7}$$

and it can be seen as a basis for the following Uzawa methods. The *classical Uzawa algorithm* [5, Chapter 10] relies on exactly solving (4.6) and an Jacobi-like iteration for (4.7). The *Arrow-Hurwicz algorithm* [5, Chapter 6] may be regarded as an inexpensive alternative to Uzawa's classical method, see below. The latter reads as follows:

Algorithm 4.2.2 (Classical Uzawa method)**.** *Given p_0, find u_{k+1}, p_{k+1} with*

$$\begin{aligned} Au_{k+1} &= f - B^T p_k \\ p_{k+1} &= p_k + \tau \left(Bu_{k+1} - g\right) \\ &= p_k + \tau \left(BA^{-1}f - g - Cp_k\right) . \end{aligned}$$

As a consequence of the Schur approach the updates for u_{k+1} and p_{k+1} in the classical Uzawa approach depend on p_k only. The Uzawa algorithm is equivalent to the method of steepest descent with $C = BA^{-1}B^T$. The classical Uzawa method converges for $\alpha < 2 \cdot \left\|BA^{-1}B^T\right\|^{-1}$.
However, for large scale and ill-posed problems the exact solution of (4.6) and (4.7) is not possible in general. At best only approximate solvers for the latter are available. Avoiding the exact solution one can consider a class of iterative methods called *inexact Uzawa algorithms*. Several modifications have been proposed, such as Jacobi-like iterations or multigrid methods to avoid the exact solution of $Au = b$, cf. [5, Chapter 6], [180] or [20, Section 8.1] for a nice overview and further references.

In general, inexact Uzawa methods can be described as

Algorithm 4.2.3 (Inexact Uzawa method)**.**

$$\hat{A}\left(u_{k+1} - u_k\right) = f - Au_k - B^T p_k \tag{4.8}$$
$$\hat{C}\left(p_{k+1} - p_k\right) = Bu_{k+1} - g , \tag{4.9}$$

where $\hat{A}$ and $\hat{C}$ are symmetric positive definite matrices and can be seen as preconditioning matrices. For $\hat{A} = A$ and $\hat{C} = \frac{1}{\tau} \cdot I$ we obtain the *classical Uzawa method*, whereas $\hat{A} = \frac{1}{\mu} \cdot I$ and $\hat{C} = \frac{1}{\tau} \cdot I$ yields the *classical Arrow-Hurwicz algorithm* [5, 142]. Convergence properties of inexact Uzawa methods have been investigated e.g. by [145, 63, 25] and more recently in a unified setting with symmetric preconditioners by Zulehner [200]. The idea of various primal-dual algorithms for saddle point problems trace back to this class of methods.

4.3 Augmented Lagrangian Methods

Historically, the augmented Lagrangian method dates back to Hestenes and Powell [99, 143]. They called the algorithm *method of multipliers*. For background references on the ALM we refer to [72], [83] and the recent book by Ito & Kunisch [107]. In the context of imaging and total variation regularization the ALM has been considered for instance in [68, 69, 198, 199, 193, 106]. The major application of the augmented Lagrangian method is to solve constrained optimization problems of the form

Problem 4.3.1 (Constrained Reconstruction Problem, Primal)**.**

$$\begin{aligned} \min_{u \in \mathcal{V}} \quad & J(u) \\ s.t. \quad & Ku = f \ , \end{aligned}$$

where $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ is a functional and $K : \mathcal{X} \to \mathcal{Y}$ is a linear operator between Banach spaces $\mathcal{X}$ and $\mathcal{Y}$. The augmented Lagrangian method is an attractive method as it breaks a constrained optimization problem into smaller unconstrained subproblems. For example you can think of J as a regularization functional and of K as a compact operator in an inverse problem. Particularly in the case of inverse problems and in the presence of noise one needs to be careful with the equality constraint in this model. Due to ill-posedness and data affected by noise, we cannot directly expect to obtain an equality. An alternative is a penalized, usually separable, reconstruction model as we will see in the following section. By introducing a Lagrange multiplier function p for the constraint in (4.3.1), the Lagrangian is given by

$$L(u; p) \ = \ J(u) \ + \ \langle p, f - Ku \rangle \ . \tag{4.10}$$

Hence, in the sense of (4.1) we can formulate a primal-dual saddle point problem for (4.3.1)

$$\min_u \max_p \; L(u;p) \; .$$

Our starting problem (4.3.1) is the primal problem based on u only. By using Legendre-Fenchel duality, see (3.1.2), we can deduce the corresponding dual problem

$$\min_p \; J^*(K^*p) \; - \; \langle f, p \rangle \; ,$$

where J^* denotes the convex conjugate (Legendre-Fenchel transform) of J. Several splitting algorithms in this chapter are based on the augmented Lagrangian. The augmented Lagrangian results from extending the Lagrangian in (4.10) by adding a quadratic penalty of a constraint in the model,

$$L^\mu(u;p) \; = \; J(u) \; + \; \langle p, f - Ku \rangle \; + \; \frac{\mu}{2} \, \|Ku - f\|_2^2 \; . \tag{4.11}$$

We apply the standard Uzawa algorithm (without preconditioning) on the augmented Lagrangian above and set the dual step size to the relaxation parameter μ. The algorithm we obtain is called method of multipliers [99, 143] or augmented Lagrangian method (ALM). For problem (4.3.1) the algorithm reads as follows:

Algorithm 4.3.2 (Augmented Lagrangian Method)**.** *Let $u^0 = 0$ and $p^0 = 0$ and choose a sequence $(\mu^k)_{k\in\mathbb{N}}$ of positive parameters. For $k = 0, 1, 2, \ldots$ compute*

$$\begin{cases} u^{k+1} & = \arg\min_u \; L^{\mu^k}(u; p^k) \\ & = \arg\min_u \left\{ \; J(u) \; + \; \left\langle p^k, f - Ku \right\rangle \; + \; \dfrac{\mu^k}{2} \, \|Ku - f\|_2^2 \; \right\} \\ p^{k+1} & = p^k + \mu^k \, (f - Ku^{k+1}) \; . \end{cases}$$

Since $f - Ku^{k+1} \in \partial_p L^{\mu^k}(u^{k+1}; p)$, the augmented Lagrangian method is nothing else then the classical Uzawa algorithm 4.9 applied to the augmented Lagrangian, where the step size τ is set to the relaxation parameter μ. In other words, the Uzawa algorithm does not necessarily assume the dual step size to be the same as the relaxation parameter μ. Concerning augmented Lagrangian algorithms it is common to differentiate between stationary and non-stationary methods, depending on whether the parameters μ^k are chosen constant or variable. Recently, there have been some publications (see e.g. [74],[165] or [75]) revealing the equivalence relation between the ALM and iterative primal Bregman distance regularization with an L^2 data fidelity. This can be verified in the following way: Without affecting minimizers,

the u update in the ALM algorithm 4.3.2 can be written as:

$$\begin{aligned} u^{k+1} &= \arg\min_u \left\{ \frac{1}{2}\|Ku - f\|_2^2 + \frac{1}{\mu^k}\left(J(u) - \langle K^* p^k, u\rangle\right) \right\} \\ &= \arg\min_u \left\{ \frac{1}{2}\|Ku - f\|_2^2 + \frac{1}{\mu^k} D_J^{K^* p^k}(u, u^k) \right\} , \end{aligned}$$

which is the primal Bregman algorithm 3.3.3 with $\alpha := \frac{1}{\mu^k}$. The update of the dual variables coincides with the update in the Bregman algorithm if you consider the subgradient $K^* p^k$ in the Bregman distance.

Using Rockafellar's duality concepts [156] the ALM can equivalently be reformulated as a proximal point method applied to the dual. In comparison to methods without the augmentation of the Lagrangian, the ALM can achieve faster convergence, but one has to be careful with choosing the parameters. In [83] Glowinski et al. proved the superlinear convergence for $\mu \nearrow \infty$, but obviously this can cause instabilities due to ill-posedness if μ is too large.

4.4 The Splitting Zoo

Recently several works in literature used convex splitting algorithms for various imaging and image processing tasks [198, 199, 69, 68, 165, 193]. In the three previous sections we introduced the main ingredients for most of the methods in the "splitting zoo": saddle point problems, Uzawa approaches with preconditioning and augmented Lagrangian methods. The obvious advance of splitting methods, is that they rely on efficient solvers for separated problems for primal and dual functions. Often these subproblems are well understood and can be implemented very easily as a combination of existing algorithms.

In the following we will take a prototype of a separable variational problem to explain fundamental convex splitting ideas to get an overview of the zoo. Throughout this work we will use splitting and preconditioning techniques for different imaging and image processing tasks. For example we will use forward-backward splitting schemes for 3D Poisson TV image reconstruction, split Bregman techniques for 3D optical flow TV computation or augmented Lagrangian and inexact Uzawa techniques (preconditioned Bregmanized operator splitting) for 4D image reconstruction with total variation and optimal transport.

Separable Problems:
A variational problem is separable if we are able to decouple, respectively split, the objective functional by adding (artificial) constraints. As a result one usually

obtains several terms in the objective functional dependent on different primal variables. Common prototypes for separable problems are variational problems where the addends are dependent on certain operators. For example in regularized inverse problems usually data fidelities H are dependent on compact operators K or regularization terms are dependent on differential operators, e.g. $A = \nabla$. Let us assume we have the following separable variational problem:

Problem 4.4.1 (Primal Separable Problem)**.**

$$\min_u \{ H(u) + J(Au) \} .$$

Here you can think of a reconstruction problem with a data fidelity H, e.g. an L^2 data fidelity and a regularization functional J, e.g. TV, which depends on the differential operator $A := \nabla$. The Problem 4.4.1 is the primal problem as it depends on u only. By using the convex conjugate (Legendre-Fenchel transform) of J and H, i.e.

$$J^*(p) = \sup_v \langle p, v \rangle - J(v) , \qquad H^*(q) = \sup_w \langle q, w \rangle - H(q) , \tag{4.12}$$

we can rewrite this problem as a purely dual problem using Fenchel's duality Theorem 3.1.8,

Problem 4.4.2 (Dual Problem (D))**.**

$$\max_p \{ -J^*(p) - H^*(-A^*(p)) \}$$

If J is a closed proper convex functional we have $J^{**} = J$. Hence we obtain

$$J(Au) = \sup_p \langle p, Au \rangle - J^*(p) .$$

Thus, a primal-dual saddle point formulation of our separable Problem 4.4.1 is given by

Problem 4.4.3 (Primal-Dual Saddle Point Problem)**.**

$$\min_u \sup_p \{ H(u) + \langle p, Au \rangle - J^*(p) \} . \tag{4.13}$$

By introducing an (artificial) constraint $Au = z$, we can split the functional into two parts dependent on different primal variables.

Problem 4.4.4 (Primal Decoupled Problem)**.**

$$\min_{u,z} \{H(u) + J(z)\} \qquad s.t. \quad Au = z \qquad \textit{(decoupled)}$$

By introducing a Lagrange multiplier function p for the constraint, the Lagrangian is defined as

$$L(u, z; p) = H(u) + J(z) + \langle p, z - Au \rangle \ .$$

Now we have two primal variables and one dual variable. Hence, in the decoupled case we obtain a new saddle point formulation for Problem 4.4.4, which is given by

Problem 4.4.5 (Decoupled Primal-Dual Saddle Point Problem, (SP))**.**

$$\min_{u,z} \sup_{p} L(u, z; p) \ .$$

By applying Legendre-Fenchel duality once more, we can verify that the dual problem corresponding to Problem 4.4.4 simply is the Problem 4.4.2 above.

By adding a quadratic penalty concerning the equality constraint $Au = z$ to the Lagrangian, we obtain the following augmented Lagrangian for the decoupled Problem 4.4.4,

$$L^{\mu}(u, z; p) = H(u) + J(z) + \langle p, z - Au \rangle + \frac{\mu}{2} \|Au - z\|_2^2 \ . \tag{4.14}$$

Consequently, we can apply the augmented Lagrangian method, see Algorithm 4.3.2, to the decoupled problem, which yields the following algorithm

Algorithm 4.4.6 (Augmented Lagrangian method for Problem 4.4.4)**.**

$$\begin{cases} (u^{k+1}, z^{k+1}) &= \arg\min_{u,z} L^{\mu}(u, z; p^k) \\ &= \arg\min_{u,z} \left\{ H(u) + J(z) + \langle p^k, z - Au \rangle + \frac{\mu}{2} \|Au - z\|_2^2 \right\} \\ p^{k+1} &= p^k + \mu \left(z^{k+1} - Au^{k+1} \right) \ . \end{cases}$$

Here we have to compute a solution for the primal variables in u and z simultaneously. Since we are interested in simple and efficient sub-steps, we can split the first update by alternatively minimizing due to u and z. We started with a decoupled problem regarding u and z. If we make use of the splitting structure in different ways, then various members of the splitting zoo can be deduced. In the following sections we will present some useful representatives namely AMA, FBS, ADMM, DRS, Split Bregman and inexact Uzawa approaches (BOS).

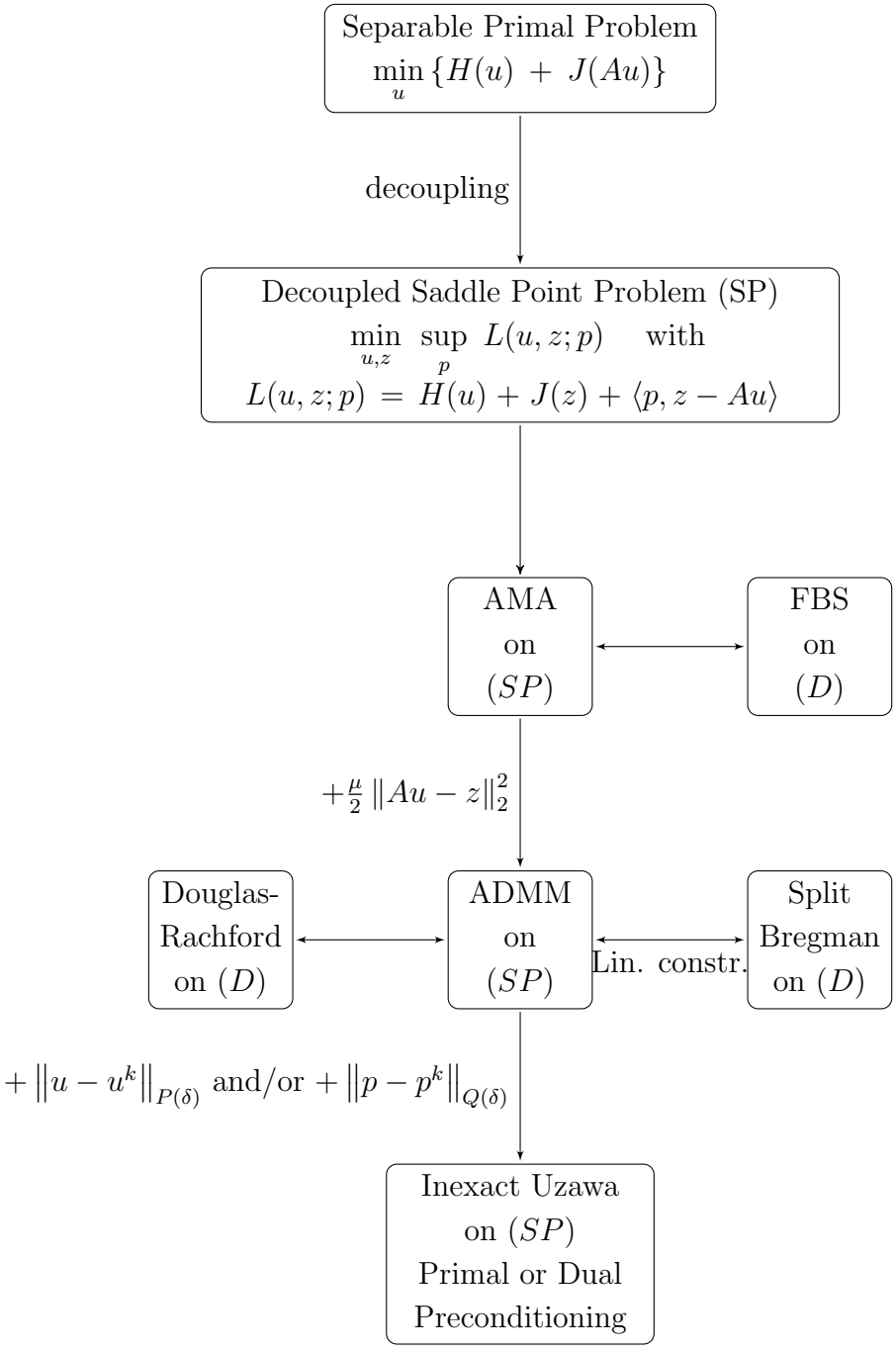


Figure 4.1: Overview of splitting methods for separable variational problems

4.4.1 Forward-Backward Splitting and AMA

We consider the decoupled primal Problem 4.4.4. In comparison to the augmented Lagrangian Algorithm 4.4.6, the main idea of the so-called *alternating minimization algorithm (AMA)* is to use the Lagrangian $L(u, z; p)$ for the update of the primal variable u, and the augmented Lagrangian $L^\mu(u, z; p)$ for the update of the other primal variable z, i.e.

Algorithm 4.4.7 (Alternating minimization algorithm (AMA) for Problem 4.4.4)**.**

$$\left\{ \begin{array}{rcl} u^{k+1} & = & \arg\min\limits_u L(u, z^k; p^k) \quad = \arg\min\limits_u \left\{ H(u) - \langle p^k, Au \rangle \right\} \\ z^{k+1} & = & \arg\min\limits_z L^\mu(u^{k+1}, z; p^k) = \arg\min\limits_z \left\{ J(z) + \langle p^k, z \rangle + \frac{\mu}{2} \left\| z - Au^{k+1} \right\|_2^2 \right\} \\ p^{k+1} & = & p^k + \mu \left(z^{k+1} - Au^{k+1} \right) . \end{array} \right.$$

This algorithm goes back to Tseng [176]. He verified that this Lagrangian based splitting algorithm is equivalent to the dual application of the so-called *forward backward splitting* (FBS) algorithm. The latter splitting technique traces back to [119] and [138], and has often be used for decomposition problems in convex optimization. We will introduce FBS for the dual Problem 4.4.2. The (dual) optimality condition for Problem 4.4.2 can be interpreted as a decomposition problem,

$$0 \in \Psi(p) + \Phi(p) , \tag{4.15}$$

with the maximal monotone operators

$$\Psi(p) := \partial J^*(p) \qquad \text{and} \qquad \Phi(p) := -A \, \partial H^*(-A^*(p)) . \tag{4.16}$$

The basic idea of the following splitting procedure (and similarly for Douglas-Rachford splitting in the following section) is to rearrange equation 4.15, such that we obtain a fixed point equation $p = T(p)$ and a related converging Picard iteration. For $\mu > 0$, equation 4.15 is equivalent to

$$(I - \mu\Psi)p \in (I + \mu\Phi)(p) ,$$

which can be written as

$$p \in (I + \mu\Phi)^{-1}(I - \mu\Psi)(p) .$$

With the notation of the resolvent operator $R_\Phi^\mu := (I + \mu\Phi)^{-1}$, we can write the fixed point equation as

$$p = T(p) := R_\Phi^\mu (I - \mu\Psi)(p) . \tag{4.17}$$

Then the corresponding Picard iteration is the FBS algorithm and reads as follows

$$p^{k+1} \;=\; R_{\Phi}^{\mu}(I - \mu\Psi)(p^k) \;.$$

We can rewrite this closed form of FBS as a two-step algorithm,

Algorithm 4.4.8 (Forward-Backward Splitting (FBS))**.**

$$\begin{cases} \dfrac{p^{k+\frac{1}{2}} - p^k}{\mu} & + \;\; \Psi(p^k) \;=\; 0 \\ \dfrac{p^{k+1} - p^{k+\frac{1}{2}}}{\mu} & + \;\; \Phi(p^{k+1}) \;=\; 0 \;, \end{cases}$$

which explains the name FBS. The first step is a forward step on Ψ, whereas the second step is a backward step on Φ.

The main advantage of AMA and equivalently of FBS applied to the dual, is the simplicity of the first iteration step. We only have to apply operators. Particularly, the use of the simple Lagrangian in the first step of AMA avoids the inversion of terms including A^*A if you take a look at the optimality condition.

However, to apply AMA, respectively FBS, either H or J must be strictly convex. Assuming that H is strictly convex with modulus $m > 0$, Tseng [176] showed that $\left\{u^k, z^k\right\}$ converges to a solution of the primal Problem 4.4.4 and $\left\{p^k\right\}$ converges to a solution of the dual Problem 4.4.2, if the time dependent step size μ^k satisfies the following restriction

$$\epsilon \;\leq\; \mu^k \;\leq\; \frac{4m}{\|A\|^2} \;-\; \epsilon \;, \qquad \epsilon \;\in\; (0, \frac{2m}{\|A\|^2}] \;.$$

Forward backward splitting can efficiently be applied to the dual problem and has successfully been used for many problems, cf. Chambolle's gradient descent reprojection algorithm [43], the Bermudez-Moreno algorithm [9] or the iterative thresholding algorithm FPC [93, 94]. In the following chapter on 3D imaging we will use a modified forward-backward splitting scheme for image reconstruction with Poisson noise and total variation regularization applied to a primal formulation.

4.4.2 ADMM and DRS and Split Bregman

In this subsection we still consider the decoupled primal Problem 4.4.4. Taking into account the augmented Lagrangian Algorithm 4.4.6, we can take advantage of the separable structure by minimizing alternatively between the separated primal variables and the corresponding dual variables. This strategy is called *alternating direction method of multipliers* (ADMM), which has been introduced by Gabay in [76].

Algorithm 4.4.9 (Alternating direction method of multipliers, ADMM)**.**

$$\left\{\begin{array}{rcl} u^{k+1} & = & \arg\min\limits_{u} L^{\mu}(u, z^k; p^k) \quad = \arg\min\limits_{u} \left\{ H(u) \; - \; \langle p^k, Au\rangle \; + \; \dfrac{\mu}{2} \left\| Au - z^k \right\|_2^2 \right\} \\ z^{k+1} & = & \arg\min\limits_{z} L^{\mu}(u^{k+1}, z; p^k) \; = \arg\min\limits_{z} \left\{ J(z) \; + \; \langle p^k, z\rangle \; + \; \dfrac{\mu}{2} \left\| z - Au^{k+1} \right\|_2^2 \right\} \\ p^{k+1} & = & p^k \; + \; \mu \left(z^{k+1} - Au^{k+1} \right) \; . \end{array}\right.$$

In comparison to AMA we have the same steps except replacing the Lagrangian in the update of the first primal variable with the augmented Lagrangian. Without affecting minimizers, Algorithm (4.4.9) can be rewritten as

$$\left\{\begin{array}{rcl} u^{k+1} & = & \arg\min\limits_{u} \left\{ H(u) \; + \; \dfrac{\mu}{2} \left\| Au - z^k - \dfrac{1}{\mu} p^k \right\|_2^2 \right\} \\ z^{k+1} & = & \arg\min\limits_{z} \left\{ J(z) \; + \; \dfrac{\mu}{2} \left\| z - (Au^{k+1} - \dfrac{1}{\mu} p^k) \right\|_2^2 \right\} \\ p^{k+1} & = & p^k \; + \; \mu \left(z^{k+1} - Au^{k+1} \right) \; , \end{array}\right.$$

because the scalar products can be transfered to the L^2 norms. The ADMM Algorithm 4.4.9 is equivalent to the so-called Douglas-Rachford Splitting (DRS) technique applied to the dual Problem 4.4.2. DRS traces back to a work of Douglas and Rachford from 1956 [58]. Connections between ADMM and DRS were explored by Glowinki and Le Tallec [83], Gabay [76] and by Lions and Mercier [119]. The equivalence of ADMM and the proximal point method was studied in [119], too.

The optimality system of the dual Problem 4.4.2 reads as follows

$$0 \in \Psi(p) \; + \; \Phi(p) \; , \tag{4.18}$$

with the maximal monotone operators defined in 4.16. This is a decomposition problem in convex optimization and can be interpreted as the steady state of a parabolic problem. By formally applying Douglas-Rachford Splitting (DRS) to (4.18), we obtain a two-step algorithm

$$\left\{\begin{array}{l} \dfrac{p^{k+\frac{1}{2}} - p^k}{\mu} \; + \; \Psi(p^{k+\frac{1}{2}}) \; + \; \Phi(p^k) \; = \; 0 \\ \dfrac{p^{k+1} - p^k}{\mu} \; + \; \Psi(p^{k+\frac{1}{2}}) \; + \; \Phi(p^{k+1}) \; = \; 0 \; , \end{array}\right. \tag{4.19}$$

where μ can be interpreted as a step size. Following [83] and [60], one can show that ADMM satisfies scheme (4.19).

In the following we will derive the DRS scheme in (4.19) by aiming fixed point equations and Picard iterations, similarly to the derivation of forward backward

splitting (FBS). We start with the fixed point relation, we used for FBS in (4.17), with the resolvent $R_\Phi^\mu := (I + \mu\Phi)^{-1}$ and $\mu > 0$,

$$p \in R_\Phi^\mu(I - \mu\Psi)(p) ,$$

which is equivalent to (4.18). By multiplying the equation with $(I + \mu\Psi)$ we obtain

$$(I + \mu\Psi)(p) \in R_\Phi^\mu(I - \mu\Psi)(p) + \mu\Psi(p) ,$$

which again is equivalent to

$$p \in R_\Psi^\mu \left(R_\Phi^\mu(I - \mu\Psi) + \mu\Psi\right)(p) . \tag{4.20}$$

By introducing a new function t defined as

$$p = R_\Psi^\mu(t) ,$$

and with the identities

$$R_\Psi^\mu - \mu\Psi(R_\Psi^\mu) = 2R_\Psi^\mu - I \quad \text{and} \quad I - \mu\Psi = 2 - (I + \mu\Psi) ,$$

the equation in (4.20) yields

$$t \in R_\Phi^\mu(2p - t) + t - p .$$

Consequently, if Ψ is single-valued, then the DRS scheme reads as follows:

Algorithm 4.4.10 (Douglas-Rachford Splitting (DRS))**.**

$$\begin{cases} t^{k+1} = R_\Phi^\mu(2p^k - t^k) + t^k - p^k \\ p^{k+1} = R_\Psi^\mu(t^{k+1}) . \end{cases}$$

If there exists a solution of the dual problem, then for any initial values t^0 and p^0 and for any step size $\mu > 0$, this algorithm converges weakly to a solution. Concerning a proof, cf. Lions and Mercier [119] or Combettes [47].

For a single-valued Ψ, the fixed point iteration regarding (4.20),

$$p^{k+1} \in R_\Psi^\mu \left(R_\Phi^\mu(I - \mu\Psi) + \mu\Psi\right)(p^k) ,$$

is a Picard iteration and represents a closed form of the DRS scheme. With the substitution

$$p^{k+\frac{1}{2}} := p^{k+1} - \mu\Psi(p^k) + \mu\Psi p^{k+1} ,$$

we can write this iteration as a two-step iteration scheme:

$$\begin{cases} \dfrac{p^{k+\frac{1}{2}} - p^k}{\mu} & + \ \Phi(p^{k+\frac{1}{2}}) \ + \ \Psi(p^k) \ = \ 0 \\ \dfrac{p^{k+1} - p^{k+\frac{1}{2}}}{\mu} & - \ \Psi(p^k) \ + \ \Psi(p^{k+1}) \ = \ 0 \ , \end{cases}$$

which is equivalent to (4.19) and which reveals the structural differences compared to the FBS Algorithm 4.4.8. Douglas-Rachford Splitting is equivalent to the so-called Split Bregman algorithm if the constraints are linear. This connection has been established by Setzer [165, 166]. The connection between Split Bregman and ADMM has been made by Esser [68, 69].

Example:
In the case of TV denoising we have the primal decoupled Problem 4.4.4 with

$$H(u) := \|u - f\|_2^2 \ , \quad \text{and} \quad J(z) := \alpha \, \|z\|_1 \ , \tag{4.21}$$

with the operator $A := \nabla$. Hence, the application of ADMM to this problem is equivalent to the Split Bregman method proposed in [84]. The Split Bregman algorithm can be realized very efficiently via shrinkage and DCT, respectively FFT inversion:

$$\begin{cases} u^{k+1} \ = \ (1 \ - \ \mu\Delta)^{-1} \left(f \ + \ \mu z^k \ + \ p^k \right) & \\ z^{k+1} \ = \ S(\nabla u^{k+1} - \frac{1}{\mu} p^k \, , \, \frac{\alpha}{\mu}) & \text{(Shrinkage)} \\ p^{k+1} \ = \ p^k \ + \ \mu \left(z^{k+1} - \nabla u^{k+1} \right) \ . & \end{cases}$$

In Section 8.5 on optical flow computation, we will use the Split Bregman idea, respectively ADMM or DRS, for computing 3D optical flow with total variation.

The quadratic relaxation terms of the form $\frac{\mu}{2} \|z - Au\|_2^2$ in the ADMM algorithm result from the augmentation of the Lagrangian. These additional penalty terms can improve the convergence behavior of resulting splitting schemes. However, the primal variables in augmented Lagrangian based splitting schemes are coupled in a way that can make sub-steps computationally expensive. If A is diagonal or diagonalizable this may not be a problem since we can simply invert operators by DCT or FFT, similarly to the TV denoising example from above. However, for more complicated operators A concerning inverse problems, e.g. in PET imaging, this is a difficulty.

There are at least two possibilities to overcome the difficulty resulting from the augmentation in the corresponding primal update. On the one hand, we can simply take the Lagrangian without the additional quadratic penalty. This leads to the

alternating minimization algorithm (AMA), respectively to forward-backward splitting (FBS) for the dual problem, we presented in the previous section. On the other hand, we can add suitable preconditioning terms relating consecutive primal iterates in a specific way. This leads to the idea of Bregmanized operator splitting, or more general, to inexact Uzawa algorithms for saddle point problems we introduced in Section 4.2. In the following subsections we will study this issue further in detail.

4.4.3 Bregmanized Operator Splitting and Inexact Uzawa

In the first part of this subsection we will introduce Bregmanized operator splitting (BOS) for the primal constrained reconstruction Problem 4.3.1, which reads as follows

$$\begin{aligned} \min_{u\in\mathcal{V}} \quad & J(u) \\ \text{s.t.} \quad & Ku = f \ . \end{aligned}$$

In Section 3.2, in Algorithm 3.3.4, we presented the primal Bregman iteration,

$$\left\{ \begin{array}{rcl} u^{k+1} & = & \arg\min\limits_{u} \left\{ \frac{1}{2} \left\| Ku - f^k \right\|_2^2 \ + \ J(u) \right\} \\ f^{k+1} & = & f^k \ + \ f - Ku^{k+1} \ , \end{array} \right.$$

which, in the case of L^2 data fidelities, turns out to be equivalent to the augmented Lagrangian method regarding the constrained reconstruction problem above, see Algorithm 4.3.2.
Then the Bregmanized operator splitting technique introduced in [198] results from 4.22 by applying forward-backward splitting (see Algorithm 4.4.8) to the first minimization problem concerning u. Restricting to one inner iteration and using the step-size δ, the BOS algorithm reads as follows:

Algorithm 4.4.11 (Bregmanized Operator Splitting (BOS) for Problem 4.3.1)**.**

$$\left\{ \begin{array}{rcl} u^{k+\frac{1}{2}} & = & u^k - \delta K^* \left(Ku^k - f^k \right) \\ u^{k+1} & = & \arg\min\limits_{u} \left\{ \frac{1}{2\delta} \left\| u - u^{k+\frac{1}{2}} \right\|_2^2 \ + \ J(u) \right\} \\ f^{k+1} & = & f^k \ + \ f - Ku^{k+1} \ , \end{array} \right.$$

which is equivalent to

$$\left\{ \begin{array}{rcl} u^{k+1} & = & \arg\min\limits_{u} \left\{ \frac{1}{2\delta} \left\| u - \left(\left(1 - \delta K^* K \right) u^k + \delta K^* f^k \right) \right\|_2^2 \ + \ J(u) \right\} \\ f^{k+1} & = & f^k \ + \ f - Ku^{k+1} \ . \end{array} \right.$$

It turns out, that BOS can be interpreted as a special case of the inexact Uzawa method we introduced in Section 4.2, also see [198].
In the sense of Lagrangian methods, the inexact Uzawa approach regarding the constrained reconstruction Problem above, reads as follows

Algorithm 4.4.12 (Inexact Uzawa Algorithm for Problem 4.3.1)**.**

$$\begin{cases} u^{k+1} &= \arg\min\limits_{u} \left\{ J(u) \;+\; \langle f - Ku, f^k \rangle \;+\; \|Ku - f\|_2^2 \;+\; \left\|u - u^k\right\|_{\hat{A}}^2 \right\} \\ \hat{C}\, f^{k+1} &= \hat{C}\, f^k \;+\; f - Ku^{k+1} \;, \end{cases}$$

with the preconditioning term

$$\left\|u - u^k\right\|_{\hat{A}}^2 = \left\langle u - u^k, \hat{A}(u - u^k) \right\rangle$$

and the preconditioning operators $\hat{A}$ and $\hat{C}$, we already introduced in Algorithm 4.2.3. By carefully choosing these operators, we are able to "control" subsequent primal respectively dual iterates in a specific way.

If we choose $\hat{A} := \frac{1}{\delta} - K^*K$ and $\hat{C} := I$, this algorithm coincides with Bregmanized operator splitting.

In Chapter 10 we will apply the general inexact Uzawa strategy with preconditioning to a 4D TV image reconstruction model with optimal transport to obtain an efficient splitting scheme for a 4D-TV image reconstruction model. As a consequence, we are going to build sub-steps, which consist of thresholding and DCT inversions only.

Example: Inexact Uzawa for $L^2 - TV$ Reconstruction
In the following we will present Bregmanized operator splitting (BOS), respectively inexact Uzawa methods applied to image reconstruction problems with a *separable* structure. This example can be seen as a motivation for the splitting algorithms we will use at the end of this thesis.

Let us assume the following variational problem for image reconstruction,

$$\min_u \left\{ \frac{1}{2} \|Ku - f\|_2^2 \;+\; \alpha\, TV(u) \right\} \;,$$

with an L^2 data fidelity corresponding to an additive Gaussian noise modeling and total variation regularization. Similarly to the previous subsections, the data fidelity and the reconstruction term can be decoupled by adding constraints.

Problem 4.4.13 (Decoupled Reconstruction Problem $L^2 - TV$)**.**

$$\min_{u,z} \left\{ \frac{1}{2} \|Ku - f\|_2^2 \;+\; \alpha \;\|z\|_1 \right\} \qquad s.t. \;\; z = \nabla u \quad \textit{(decoupled)} \;.$$

This is the decoupled primal Problem 4.4.4 with

$$H(\cdot) := \frac{1}{2}\,\|K(\cdot) - f\|_2^2 \quad \text{and} \quad J(\cdot) := \alpha\ \|\cdot\|_1\ .$$

The augmented Lagrangian regarding Problem 4.4.13 reads as follows,

$$L(u,z;p)\ :=\ \frac{1}{2}\,\|Ku-f\|_2^2\ +\ \alpha\ \|z\|_1\ +\ \langle p, \nabla u - z\rangle\ +\ \frac{\mu}{2}\,\|\nabla u - z\|_2^2\ .$$

With the additional preconditioning term

$$\left\|u-u^k\right\|_{P_\delta}^2 = \left\langle u-u^k, P_\delta(u-u^k)\right\rangle\ ,$$

we obtain

Algorithm 4.4.14 (Inexact Uzawa Algorithm)**.**
$u^0 = 0, z^0 = 0, y^0 = 0$

$$\left\{\begin{array}{rcl} u^{k+1} & = & \arg\min\limits_u \left\{ \frac{1}{2}\,\|Ku-f\|_2^2\ +\ \langle p^k, \nabla u\rangle\ +\ \frac{\mu}{2}\,\left\|\nabla u - z^k\right\|_2^2\ +\ \frac{1}{2}\,\left\|u-u^k\right\|_{P_\delta}^2 \right\} \\ z^{k+1} & = & \arg\min\limits_z \left\{ \alpha\ \|z\|_1\ -\ \langle p^k, z\rangle\ +\ \frac{\mu}{2}\,\left\|\nabla u^{k+1} - z\right\|_2^2\ +\ \frac{1}{2}\,\left\|z-z^k\right\|_Q^2 \right\} \\ \hat{C}\,p^{k+1} & = & \hat{C}\,p^k\ +\ \mu\left(\nabla u^{k+1} - z^{k+1}\right)\ . \end{array}\right.$$

For the semi-implicit Bregmanized operator splitting technique introduced in [198], we set

$$Q := 0 \quad \text{and} \quad \hat{C} := \frac{1}{\tau}\ .$$

$$\left\{\begin{array}{rcl} u^{k+1} & = & \arg\min\limits_u \left\{ \frac{1}{2}\,\|Ku-f\|_2^2\ +\ \frac{\mu}{2}\,\left\|\nabla u - z^k + \frac{1}{\mu}p^k\right\|_2^2\ +\ \frac{1}{2}\,\left\|u-u^k\right\|_{P_\delta}^2 \right\} \\ z^{k+1} & = & \arg\min\limits_z \left\{ \alpha\ \|z\|_1\ +\ \frac{\mu}{2}\,\left\|z - \left(\nabla u^{k+1} + \frac{1}{\mu}p^k\right)\right\|_2^2 \right\} \\ p^{k+1} & = & p^k\ +\ \tau\left(\nabla u^{k+1} - z^{k+1}\right)\ . \end{array}\right.$$

Numerical solution of the primal variational problems:

$$0\ =\ K^*(Ku^{k+1}-f) - \mu\,\nabla\cdot\left(\nabla u^{k+1} - z^k + \frac{1}{\mu}p^k\right)\ +\ P_\delta(u^{k+1}-u^k)$$

which implies

$$u^{k+1}\ =\ (K^*K - \mu\,\Delta + P_\delta)^{-1}\left(K^*f + \nabla\cdot\left(p^k - \mu\,z^k\right) + P_\delta u^k\right)\ .$$

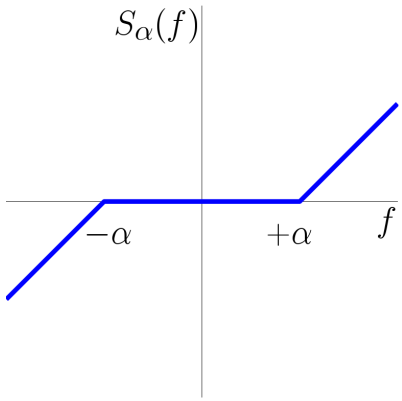


Figure 4.2: Visualization of soft-shrinkage

With the preconditioning operator $P_\delta := \frac{1}{\delta} - K^*K$ we obtain a simplified update for u,

$$u^{k+1} = \left(\frac{1}{\delta} - \mu\,\Delta\right)^{-1}\left(K^*f + \nabla\cdot\left(p^k - \mu\, z^k\right) + \left(\frac{1}{\delta} - K^*K\right)u^k\right) ,$$

which can be inverted efficiently via DCT or FFT. The second step of the algorithm, the z update, is based on a functional with an L^2 fitting term and a sparsity regularization term. Considering the optimality condition of the functional concerning z, we obtain a solution via thresholding, i.e. via the simple shrinkage formula

$$z^{k+1} = S\left(\nabla u^{k+1} + \frac{1}{\mu}p^k\,,\ \frac{\alpha}{\mu}\right) .$$

Soft-Shrinkage Formulas:
In the following we will specify the isotropic and anisotropic soft-shrinkage formulas, which are given by simple thresholding. The thresholding idea is illustrated in Figure 4.4.3. Throughout this thesis we will use these formulas to solve denoising (sub-)problems with sparsity regularization as above.

$$\begin{aligned} z = S(f,\alpha) &= f - \alpha\,\Pi_{\{g\,|\,|g|_{l^1}\le 1\}}\left(\frac{f}{\alpha}\right) \\ &= \begin{cases} f - \alpha\, sign(f)\,, & \text{if } |f| > \alpha \\ 0\,, & \text{if } |f| \le \alpha \end{cases} = \begin{cases} f-\alpha\,, & \text{if } f > \alpha \\ 0\,, & \text{if } -\alpha \le f \le \alpha \\ f+\alpha\,, & \text{if } f < \alpha \end{cases} \\ &= sign(f)\cdot max(|f|-\alpha, 0)\ . \end{aligned}$$

This is the anisotropic case of soft-shrinkage. In the case where z and f are vector fields, i.e. $z : \Omega \to \mathbb{R}^d$ with $d \ge 2$ we can choose different norms analog to the definition of total variation. Hence for isotropic soft-shrinkage we obtain the following

shrinkage formula

$$\text{For } i = 1, .., d: \qquad z_i = S(f, \alpha) = f - \alpha \, \Pi_{\{g \mid |g|_{l^2} \leq 1\}} \left(\frac{f}{\alpha} \right)$$
$$= \frac{f_i}{|f|_{l^2}} \cdot max(|f|_{l^2} - \alpha, 0) \; .$$

5 3D Imaging

5.1 Introduction

Image reconstruction is a fundamental problem in several areas of applied sciences, such as medical imaging, optical microscopy or astronomy. An interesting example is positron emission tomography (PET), a biomedical imaging technique in nuclear medicine that generates images of living organism by visualizing weak radioactively marked pharmaceuticals, so-called tracers. Due to the possibility of measuring temporal tracer uptake (from list-mode data), this modality is particularly suitable for investigating physiological and biochemical processes. Another application of image reconstruction is fluorescence microscopy, which represents an important technique for investigating biological living cells at nanoscales. In this type of applications image reconstruction arises in terms of deconvolution problems, where the undesired blurring effects being caused by diffraction of light.

Mathematically, image reconstruction in those applications can be formulated as a linear inverse and ill-posed problem. Typically, in such problems one has to deal with Fredholm integral equations of the first kind, or more general

$$\bar{f} = \bar{K}\bar{u}$$

with a compact linear operator $\bar{K}$, exact data $\bar{f}$ and the desired exact image $\bar{u}$. Unfortunately, in practice only noisy versions f and K of $\bar{f}$ and $\bar{K}$ are available and an approximate solution u of $\bar{u}$ from

$$f = Ku \tag{5.1}$$

is wanted. The computation of u by direct inversion of K is not reasonable since (5.1) is ill-posed. In this case, regularization techniques are required to enforce stability during the inversion process and to compute useful reconstructions.

A commonly used idea to realize regularization techniques with statistical motivation is the Bayesian model, using the posterior probability density $p(u|f)$, given according to Bayes formula

$$p(u|f) \;\sim\; p(f|u)\, p(u)\;. \tag{5.2}$$

The computationally interesting Bayesian approach is the maximum a-posteriori (MAP) likelihood estimation, which consists of computing an estimate u of the unknown object by maximizing the a-posteriori probability density $p(u|f)$. If the measurements f are available, the density $p(u|f)$ is denoted as the a-posteriori likelihood function, just depending on u. The Bayesian approach (5.2) has the advantage that it allows to incorporate additional prior information about u via the a-priori probability density $p(u)$ into the reconstruction process. The most frequently used a-priori densities are Gibbs functions

$$p(u) \;\sim\; e^{-\alpha J(u)}\;, \tag{5.3}$$

where α denotes a positive parameter and J a convex energy functional. Typical examples for probability densities $p(f|u)$ in (5.2) are exponentially distributed raw data f. In the canonical case of additive white Gaussian noise with expected value 0 and variance σ^2 one finds

$$p(f|u) \;\sim\; e^{-\frac{1}{2\sigma^2}\|Ku-f\|^2_{L^2(\Sigma)}}\;,$$

and the minimization of the negative log-likelihood function leads to the classical Tikhonov regularization methods [22], based on minimizing a functional of the form

$$\min_{u\geq 0}\; \frac{1}{2}\,\|Ku-f\|^2_{L^2(\Sigma)} \;+\; \alpha\, J(u)\;. \tag{5.4}$$

The first, so-called data fidelity term, penalizes the deviation from the equality in (5.1), while $J(u)$ is a convex regularization term penalizing deviations from a certain ideal structure (smoothness) of the solution. If we choose $K = Id$ and the total variation (TV) regularization technique, $J(u) := |u|_{BV(\Omega)}$ as in (5.6), we obtain the well-known ROF model [158] for image denoising. An additional positivity constraint as in (5.4) is essential for common applications, since the unknown functions u usually represent densities or intensity information.

In the applications mentioned above the measured data are stochastic due to the radioactive decay of tracers in PET imaging and to laser scanning techniques in fluorescence microscopy. The random variables of the measured data in those applications are not Gaussian- but Poisson-distributed [178] with expected value given by $(Ku)_i$,

$$p(f|u) \;=\; \prod_i \frac{(Ku)_i^{f_i}}{f_i!}\, e^{-(Ku)_i}\;.$$

In this work, we will concentrate on MAP estimates for inverse problems with Poisson distributed data. The MAP estimation via the negative log-likelihood function (5.2) asymptotically leads to the following variational problem [22],

$$\min_{u \geq 0} \int_{\Sigma} (Ku - f \log Ku) \, d\mu \; + \; \alpha \, J(u) \; . \tag{5.5}$$

Up to additive terms independent of u, the data fidelity term here is the so-called Kullback-Leibler divergence (also known as cross entropy or I-divergence) between the two probability measures f and Ku. A particular complication of (5.5) compared to (5.4) is the strong nonlinearity in the data fidelity term and resulting issues in the computation of minimizers.

In the literature there are in general two classes of reconstruction methods that are used. On the one hand analytical (direct) methods and on the other hand algebraic (iterative) strategies. A classical representative for a direct method is the Fourier-based filter backprojection (FBP). Although FBP is well understood and can be computed efficiently, iterative strategies receive more and more attention in practice. The major reason is the high noise level (low signal-to-noise ratio) and the special type of statistics found in measurements of various applications, such as PET or fluorescence microscopy, which cannot be taken into account by direct methods. Thus, in this work we deal with extensions of the expectation-maximization (EM) or Richardson-Lucy algorithm [55], [122], [153], which is currently the standard iterative reconstruction method to compute (5.5) with incomplete Poisson data f [168] in the absence of regularization ($J(u) = 0$) . However, it is difficult to generalize the EM algorithm to regularized cases. The robust and accurate solutions of this problem for appropriate models of J and its analysis are one of the main contributions of this chapter.

The specific choice of the regularization functional J in (5.5) is important for the way a-priori information about the expected solution is incorporated into the reconstruction process. Smooth, in particular quadratic regularizations have attracted most attention in the past, mainly due to simplicity in analysis and computation. However, such regularization approaches always lead to blurring of the reconstructions, in particular they cannot yield reconstructions with sharp edges. Hence, recently singular regularization energies, especially those of ℓ^1 or L^1-type, have attracted strong attention. Quite recently, Bardsley proposed in [14] a theoretical framework for the regularization of Poisson likelihood estimation problems, referring to a joint work with Luttman [16], and verifies the properties of regularization schemes for different terms of J. Hence we focus on total variation (TV) regularization, which has been derived as a denoising technique in [158] and generalized to various other

imaging tasks subsequently. The exact definition of TV [1], used in work, is

$$J(u) \;:=\; |u|_{BV(\Omega)} \;=\; \sup_{\substack{g\in C_0^\infty(\Omega,R^d)\\ ||g||_\infty\leq 1}} \int_\Omega u\ \nabla\cdot g\ , \tag{5.6}$$

which is formally (true if u is sufficiently regular)

$$|u|_{BV(\Omega)} \;=\; \int_\Omega |\nabla u|\ .$$

The space of functions with bounded total variation is denoted by $BV(\Omega)$. For further properties and details of BV functions, we refer to [1], [70]. The motivation for using TV is the effective suppression of noise and the realization of almost homogeneous regions with sharp edges. These features are attractive for PET and nanoscopic imaging if the goal is to identify object shapes that are separated by sharp edges and shall be analyzed quantitatively.

Various methods have been suggested for the TV regularized variational problem (5.5), but still with limited success, e.g. [109], [136] for PET, [13], [16], [57] for deconvolution problems or [116] for denoising problems with $K = Id$. This limitation can be traced back to strong computational difficulties in the minimization of (5.5) with TV regularization and to remaining blurring effects in reconstructed images, which arises from using approximations of TV by differentiable functionals

$$J_\epsilon(u) = \int_\Omega \sqrt{|\nabla u|^2+\epsilon}\ , \qquad \epsilon>0\ .$$

In [13], Bardsley proposed an efficient computational method based on gradient projection and lagged-diffusivity where the non-negativity constraint is guaranteed via a simple projection onto the feasible set. On the other hand, the schemes suggested in [57], [109] and [136] are realized as elementary modifications of the EM algorithm, with fully explicit or semi-implicit treatment of TV in the iteration. A major disadvantage of these approaches is that the regularization parameter α needs to be chosen very small, since otherwise the positivity of solutions is not guaranteed and the EM-based algorithm cannot be continued. Due to the additional parameter dependence on ϵ these algorithms are even less robust.

In this work, we propose a robust algorithm without approximation of TV, i.e. we use (5.6) respectively a dual version. This enables us to realize cartoon reconstructions with sharp edges. We use a forward-backward splitting (FBS) approach [26], [119], [176], which can be realized by alternating classical EM steps with weighted Rudin-Osher-Fatemi (ROF) problems, which are solved analogous to the algorithm of Chambolle in [42]. The advantage of our approach is that it can be performed

equally well for large regularization parameter. Thus, it is favorably applicable for problems with a low signal-to-noise ratio. Additionally, we study the existence, uniqueness and stability of the minimization problem, prove positivity preservation of the algorithm and provide a convergence analysis for a damped FBS strategy.

It is well-known that images reconstructed with TV-based methods suffer from a loss in contrast. Hence, we suggest to extend EM-TV to an iterative regularization using Bregman iterations, which incorporates simultaneous contrast correction. The contrast improvement is realized via inverse scale space methods and Bregman iterations, introduced in [134], [34], [33]. Related to these methods, an iterative contrast correction can be implemented as a sequence of modified EM-TV problems (5.5) with

$$J(u) \;:=\; D^{\tilde{p}}_{|\cdot|_{BV(\Omega)}}(u,\tilde{u})\;, \qquad \tilde{p} \;\in\; \partial\,|\tilde{u}|_{BV(\Omega)}\;,$$

as regularization functional. The Bregman distance based on TV $D_{|\cdot|_{BV(\Omega)}}$ penalizes deviations from piecewise constant functions and does not affect the position of image edges. However, the Bregman iteration facilitates contrast improving intensity changes and enables improved reconstructions.

This chapter is organized as follows. In Section 5.2 we will recall a mathematical model for inverse problems with Poisson noise. Starting from a statistical view of the image reconstruction in form of a maximum a-posteriori (MAP) likelihood estimation based on works of [22], we will proceed to a continuous representation in terms of multidimensional variational problems. An important point in this context is the realization of a-priori knowledge via regularization functionals. As a simple special case, we will derive the well-known EM or Richardson-Lucy algorithm with positivity constraints. In Section 5.3, the EM algorithm will be combined with total variation regularization. We will deduce a robust EM-TV algorithm, realized as a two-step iteration scheme, and we will provide suitable stopping criteria. The method will be reinterpreted as a modified forward-backward splitting algorithm known from convex optimization. In Section 5.4 we will study the analysis of the EM-TV model. After proving the well-posedness of the minimization problem in terms of existence, uniqueness and stability, we will provide a convergence analysis and positivity preservation properties of the proposed EM-TV algorithm. As a consequence we obtain damping conditions to guarantee convergence of the forward-backward splitting algorithm. The numerical realization of ROF related problems, appearing in the second half step of the EM-TV algorithm, are studied in section 5.5. In the last section we will apply the primal and dual Bregman methods from Section 3.3 to the EM-TV model. In the resulting primal and dual Bregman-EM-TV methods a loss of contrast will be enhanced by iterative regularization with Bregman distances. The performance of our techniques will be illustrated by synthetic and

experimental 2D and 3D reconstructions in high-resolution fluorescence microscopy and positron emission tomography in medical imaging in the following chapter.

5.2 Modeling and EM Algorithm

5.2.1 Model for Data Acquisition

This section provides an overview of mathematical modeling essential for a reasonable formulation of inverse problems with Poisson noise. In the following we will just concentrate on the relevant aspects of the model construction and refer to the work of Bertero [22] for a detailed discussion.

An imaging system consists in general of two structural elements:

- A collection of different physical components which generate signals containing useful information of spatial properties of an object.
- A detector system that provides measurements of occurring signals and which causes in many cases the undesirable sampling and noise effects.

Hence, we assume that the raw data have following properties:

- The data are discrete and the discretization is specified by the physical configuration of the detectors. We assume that the data are given in form of a vector $f \in \mathbb{R}^N$.
- The data are realizations of random variables, since noise is a random process caused by the detector system. So, we consider the detected value f_i as a realization of a random variable F_i.

Additionally, a modeling of the imaging apparatus is necessary which describes the generation and expansion of signals during the data acquisition process. Mathematically, the aim is to find a transformation that maps the spatial distribution of an object to the signals arriving at the detectors. In this work we concentrate on problems where the transformation is a linear operator and the data acquisition process can be described by a linear equation

$$\bar{f} \;=\; \bar{K}\bar{u}\;. \tag{5.7}$$

Here, $\bar{K} : L^1(\Omega) \to L^1(\Sigma)$ is a compact linear operator with a nonclosed range that additionally preserves positivity. A typical example of (5.7) is a Fredholm integral equation of the first kind with

$$(\bar{K}u)(x) \;=\; \int_\Omega \bar{k}(x,y)u(y)\;dy\;, \qquad x \;\in\; \Sigma\;,$$

where $\bar{k}$ is a nonnegative kernel. In (5.7) the function $\bar{u}$ describes the desired exact properties of the object and $\bar{f}$ denotes the exact signals before detection. Problem statements of the type above, can be found in numerous real-life applications, such as positron emission tomography (PET) [133], [168], [189] in medical imaging, fluorescence microscopy [98], [57], astronomy [22] or radar imaging. The modeling of the data acquisition in this manner turns the problem of object reconstruction to the solution of a linear inverse problem of the form (5.7). However, as mentioned above, in practice only noisy (and discrete) versions f and K of the exact data $\bar{f}$ and operator $\bar{K}$ are available so that only an approximate solution u of $\bar{u}$ can be computed from the equation

$$f \;=\; Ku \;. \tag{5.8}$$

The operator $K : L^1(\Omega) \to \mathbb{R}^N$ here is a semi-discrete operator based on $\bar{K}$, which transforms the desired properties u, in contrary to $\bar{K}$, to the discrete raw data. Due to noise in the measurements f a model for the probability density of the noise is necessary to obtain a complete modeling for the problem of image reconstruction. In this work we concentrate on a specific non-Gaussian noise, namely the so-called Poisson noise. This type of noise appears for example in PET due to radioactive decay of tracers and due to counting of photon coincidences [178], [10], or in optical nanoscopy due to photon counts by laser sampling of an object [131], [57]. In such cases, every F_i corresponds to a Poisson random variable with an expectation value given by $(Ku)_i$, i.e.

$$F_i \text{ is Poisson-distributed with parameter } (Ku)_i \;. \tag{5.9}$$

In the following, we denote by $p(f|u)$ the conditional probability density of data f given an image u. Additionally, we make the assumption that the random variables F_i are pairwise independent and identically distributed (i.i.d.), i.e.

$$p(f|u) \;=\; \prod_{i=1}^{N} p(f_i|u) \;.$$

This assumption is in general reasonable since each random variable can be assigned to a specific detector element. Combined with (5.9), this property leads to the probability density

$$p(f|u) \;=\; \prod_{i=1}^{N} \frac{(Ku)_i^{f_i}}{f_i\,!}\, e^{-(Ku)_i} \;. \tag{5.10}$$

Hence, a complete model for the process of data generation and data acquisition is available, if the operator K and the conditional probability density $p(f;u)$ are known.

5.2.2 Statistical Problem Formulation of Image Reconstruction

Due to the compactness of inverting the operator $\bar{K}$ (5.7) is an ill-posed problem [65], [86]. Note that the problem (5.8) is not ill-posed in strong sense, because the operator K has a finite range. Nonetheless the problem is highly ill-conditioned, since K approximates $\bar{K}$, hence still some type of regularization is required to enforce stability during the inversion process and to compute useful reconstructions. A frequently used class of regularization techniques are variational methods based on the minimization of functionals of the form

$$\frac{1}{s}\,||Ku - f||^s_{L^s(\Sigma)} \;+\; \alpha\, J(u)\;, \qquad \alpha > 0\;, \quad 1 \le s < \infty\;. \tag{5.11}$$

However, from the viewpoint of statistical modeling, the functionals in (5.11) are inappropriate for problems with Poisson-distributed data, since they result from the assumption of exponentially distributed raw data $f = K\bar{u} + \eta$, where η is a vector valued random variable with statistically i.i.d. components. Typical examples are that η is Laplace-distributed ($s = 1$) or Gaussian-distributed ($s = 2$) [22].

In the following, we provide a statistical problem formulation of image reconstruction in the case of Poisson noisy raw data. Inserting the given measurements f into Bayes' formula, the density $p(u|f)$ is denoted as the a-posteriori likelihood function, which depends on u only. Now, to determine an approximation $\hat{u}$ to the unknown object $\bar{u}$, we use the maximum a-posteriori probability (MAP) estimator which maximizes the likelihood function, i.e.

$$\hat{u} \;\in\; \underset{\substack{u \in L^1(\Omega)\\ u \ge 0 \;\; a.e.}}{\arg\max}\; p(u|f)\;. \tag{5.12}$$

The positivity constraint on the solution is needed, since in typical applications the functions represent densities or intensity information.

For a detailed specification of the likelihood function in (5.12) we proceed on the assumption that a model for the process of data acquisition and data generation, in the manner of the last section, is available. For this reason, we plug the probability density for Poisson noise (5.10) and the Gibbs a-priori density concerning J (see (2.6)) in the definition of the likelihood function (5.12) and obtain the negative log-likelihood function

$$-\log p(u|f) \;=\; \sum_{i=1}^{N} \left[(Ku)_i - f_i \log(Ku)_i\right] \;+\; \alpha\, J(u) \tag{5.13}$$

in which the additive terms independent of u are neglected. At this point we will pass over from a discrete to a continuous representation of data, which corresponds

to the way events on detectors are measured. With the indicator function

$$\mathbf{1}_i(x) \;=\; \begin{cases} 1\,, & x \,\in\, M_i \\ 0\,, & \text{else} \end{cases}\,,$$

where M_i is the region of the i-th detector, we can interpret the mean values of the discrete data as

$$f_i \;=\; \int_{M_i} f\; dx \;=\; \int_{\Sigma} \chi_i f\; dx\;.$$

Thus we can rewrite the MAP estimate in (5.12) as the following continuous variational problem,

$$\hat{u} \;\in\; \underset{\substack{u \,\in\, L^1(\Omega) \\ u \,\geq\, 0 \;\; a.e.}}{\arg\min} \int_{\Sigma} (Ku - f \log Ku)\; d\mu \;+\; \alpha\, J(u) \tag{5.14}$$

with $d\mu = \sum_{i=1}^{N} \chi_i(x)\; d\lambda$, where λ denotes the Lebesgue measure.

A particular complication of (5.14) compared to (5.11) is the strong nonlinearity in the data fidelity term and resulting issues in the computation of minimizers. Finally, with respect to problem (5.14), we point out that the functional J in the Gibbs a-priori density (see (2.6)) is related to a regularization functional in the context of inverse problems, cf. (5.11). Due to the problem formulation of the image reconstruction via Bayes' theorem one refers to Bayesian regularization in this context.

5.2.3 Reconstruction Method: EM Algorithm

In the previous section we presented a statistical problem formulation for inverse problems with measured data drawn from Poisson statistics and could observe that the Bayesian approach leads to a constrained minimization problem (5.14). In this section we will give a review on a popular reconstruction algorithm for this problem, so-called expectation-maximization (EM) algorithm [55], [133], [168], which finds numerous applications, for instance in medical imaging, microscopy or astronomy. In the two latter ones, the algorithm is also known as Richardson-Lucy algorithm [122], [153]. The EM algorithm is an iterative procedure to maximize the likelihood function $p(u|f)$ in problems with incomplete data and will form a basis for our algorithms introduced later. Here, we disregard the prior knowledge and assume that any object u has the same relevance, i.e. the Gibbs a-priori density $p(u)$ is constant. For simplicity we normalize $p(u)$ such that $J(u) \equiv 0$. Hence, the problem

in (5.14) reduces to the following variational problem with a positivity constraint,

$$\min_{\substack{u \in L^1(\Omega) \\ u \geq 0 \ a.e.}} \int_\Sigma (Ku - f \log Ku) \, d\mu \ . \tag{5.15}$$

To derive the algorithm, we consider the first order optimality condition of the constrained minimization problem (5.15). Formally, the Karush-Kuhn-Tucker (KKT) conditions [100, Theorem 2.1.4] provide the existence of a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of the functional in (5.15) need to fulfill the equations

$$\begin{cases} 0 &= K^* \mathbf{1}_\Sigma \ - \ K^* \left(\dfrac{f}{Ku} \right) \ - \ \lambda \\ 0 &= \lambda u \end{cases} \tag{5.16}$$

where K^* is the adjoint operator of K and $\mathbf{1}_\Sigma \in (L^1(\Sigma))^* = L^\infty(\Sigma)$ is the characteristic function on Σ. The optimization problem (5.15) is convex, hence every function u fulfilling the equations (5.16) is a global minimum of (5.15). Multiplying the first equation in (5.16) by u and division by $K^*\mathbf{1}_\Sigma$ yields the fixed point equation

$$u \ = \ u \, \frac{K^*}{K^*\mathbf{1}_\Sigma} \left(\frac{f}{Ku} \right) \ - \ \frac{1}{K^*\mathbf{1}_\Sigma} \, \lambda \, u \ .$$

Thus, the Lagrange multiplier λ can be eliminated and we obtain a simple iteration scheme,

$$u_{k+1} \ = \ u_k \, \frac{K^*}{K^*\mathbf{1}_\Sigma} \left(\frac{f}{Ku_k} \right) \ , \tag{5.17}$$

which preserves positivity if the operator K preserves positivity and the initialization u_0 is positive. This iteration scheme is the well-known EM algorithm, respectively Richardson-Lucy algorithm. In [168], Shepp and Vardi showed that this iteration is a closed example of the EM algorithm proposed by Dempster, Laird and Rubin in [55], who presented the algorithm in a more general setup.

In the case of noise-free data f several convergence proofs of the EM algorithm to the maximum likelihood estimate (5.12), i.e. the solution of (5.15), can be found in literature [133], [151], [178], [108]. Besides, it is known that the speed of convergence of the iteration (5.17) is slow.

For noisy data f it is necessary to differentiate between discrete and continuous modeling. In the fully discrete case, i.e. if K is a matrix and u is a vector, the existence of a minimizer can be guaranteed, since the smallest singular value is bounded away from zero by a positive value. Hence, the iterates are bounded during the iteration and convergence is ensured. A further property of the iteration is a lack of smoothing, whereby the so-called "checkerboard effect" arises, i.e. single pixels

become visible in the iterates. However, if K is a general continuous operator, the convergence is not only difficult to prove, but even a divergence of the EM algorithm is possible, due to the underlying ill-posedness. This aspect can be taken as a lack of additional prior knowledge about the unknown u resulting from $J(u) \equiv 0$.

As described in [151], the EM iterates show the following typical behavior for ill-posed problems. The (metric) distance between the iterates and the solution decreases initially before it increases as the noise is amplified during the iteration process. This issue might be regulated by using appropriate stopping rules to obtain reasonable results. In [151] it is shown that certain stopping rules indeed allow stable approximations. Another possibility to considerably improve reconstruction results are regularization techniques. In the following chapter we shall discuss techniques with edge preserving properties.

5.3 EM-TV Reconstruction Method

5.3.1 Algorithm FB-EM-TV

The EM or Richardson-Lucy algorithm, discussed in Section 5.2.3, is currently the standard iterative reconstruction method for most inverse problems with incomplete Poisson data based on the linear equation (5.8). However, with the assumption $J(u) \equiv 0$, no a-priori knowledge about the expected solution are taken into account, i.e. different images have the same a-priori probability. Especially in case of measurements with low SNR, like lower tracer dose rate or tracer with short radioactive half life in case of PET examinations, the multiplicative fixed point iteration (5.17) delivers unsatisfactory and noisy results even with early termination. Therefore, we propose to integrate nonlinear variational methods into the reconstruction process to make an efficient use of a-priori information and to obtain improved results.

An interesting approach to improve the reconstructions is the EM-TV algorithm. In the classical EM algorithm, the negative log-likelihood functional (5.15) is minimized. In the EM-TV approach, we modify the functional by adding a weighted total variation (TV) term [158],

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \ \text{a.e.}}} \int_\Sigma (Ku - f \log Ku)\, d\mu \ + \ \alpha\, |u|_{BV(\Omega)} \ , \qquad \alpha > 0 \ . \tag{5.18}$$

This variational problem is exactly (5.14) with TV regularization, i.e. from the statistical point of view in Section 5.2.2, we use an a priori probability density $p(u)$ concerning $J(u) = |u|_{BV(\Omega)}$. This means that images with smaller total variation

(higher prior probability) are preferred in the minimization (5.18). The expected reconstructions are cartoon-like images, i.e. they will result in almost uniform (mean) intensities inside the different structures which are separated by sharp edges. Obviously, such an approach cannot be used for studying certain properties inside the structures in an object (which is anyway unrealistic in case of low SNR), but it is well suited for segmenting different structures and analyzing them quantitatively.

TV regularization is a popular and important concept in several fields of mathematical image processing. It has been derived as a denoising technique in [158] and generalized to various other imaging tasks subsequently. The space of functions with bounded total variation is denoted by $BV(\Omega)$ and is a Banach space equipped with the norm

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

$BV(\Omega)$ is a popular function space in image processing since it can represent discontinuous functions (related to the edges in an image) which are even preferred during the minimization of TV. For further properties and details on functions with bounded variation we refer to [1], [70], [81].

For designing the proposed alternating algorithm, we consider the first-order optimality condition of (5.18). Due to the total variation, this variational problem is not differentiable in the usual sense. However, we can extend the data fidelity term to a convex functional without changing the stationary points,

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \;\; a.e.}} D_{KL}(f, Ku) \;+\; \alpha\, |u|_{BV(\Omega)} \tag{5.19}$$

with the Kullback-Leibler (KL) functional D_{KL} [150],

$$D_{KL}(v, u) = \int_\Sigma \left(v \log \frac{v}{u} \;-\; v \;+\; u \right) \, d\mu \;,$$

such that the minimization problem (5.19) becomes convex, see section 5.4. For such problems powerful methods from convex analysis are available, see Section 3.1.

For the use of the subdifferential calculus on the functional in (5.19), note that due to the definition of $D_{KL}(f, K\,\cdot)$ on $L^1(\Omega)$, its subgradients are elements of $L^\infty(\Omega)$, while in general the subgradients of $|\cdot|_{BV(\Omega)}$ are in the larger function space $(BV(\Omega))^*$. However, we can extend TV to a convex functional on $L^1(\Omega)$ by setting $|u|_{BV(\Omega)} = \infty$ if $u \in L^1(\Omega) \setminus BV(\Omega)$. Hence, due to the continuity of the KL functional and [62, Chapter I, Proposition 5.6], we obtain the identity

$$\partial \left(D_{KL}(f, Ku) \;+\; \alpha\, |u|_{BV(\Omega)} \right) \;=\; \partial_u\, D_{KL}(f, Ku) \;+\; \alpha\, \partial\, |u|_{BV(\Omega)}$$

in $L^\infty(\Omega) \subset (BV(\Omega))^*$ for any $f \in L^1(\Omega)$. For the KL functional D_{KL} the subdifferentials are singletons and the first optimality condition of (5.19) for a positive solution u is given by

$$K^*\mathbf{1}_\Sigma - K^*\left(\frac{f}{Ku}\right) + \alpha\, p = 0\,, \qquad p \in \partial\,|u|_{BV(\Omega)}\,, \tag{5.20}$$

where K^* denotes the adjoint of K. Formally this condition is a nonlinear integrodifferential equation

$$K^*\mathbf{1}_\Sigma - K^*\left(\frac{f}{Ku}\right) - \alpha\,\nabla\cdot\left(\frac{\nabla u}{|\nabla u|}\right) = 0\,.$$

The simplest iteration scheme to compute a solution of the variational problem (5.18), respectively (5.19), is a gradient-type method in L^2, which however is not robust in case of TV and severe step size restrictions are needed since the subgradient p of TV is treated explicitely. A better idea is to use an iteration scheme which evaluates the nonlocal term (including the operator K) in (5.20) at the last iterate u_k and the local term (including the subgradient of TV) at the new iterate u_{k+1}, i.e.

$$1 - \frac{K^*}{K^*\mathbf{1}_\Sigma}\left(\frac{f}{Ku_k}\right) + \alpha\,\frac{1}{K^*\mathbf{1}_\Sigma}\,p_{k+1} = 0\,, \qquad p_{k+1} \in \partial\,|u_{k+1}|_{BV(\Omega)}\,, \tag{5.21}$$

with an additional division of (5.20) by $K^*\mathbf{1}_\Sigma$. In this iteration, the new iterate u_{k+1} appears only as a point of reference for the subdifferential of $|\cdot|_{BV(\Omega)}$. This is a considerable drawback since u_{k+1} cannot be determined from (5.21), due to the missing of an one-to-one relation between subgradients and primal variable u. In addition, such iteration schemes cannot guarantee preservation of positivity. Hence, we obtain an improved method if we approximate also the constant term 1 in (5.21) by $\frac{u_{k+1}}{u_k}$ such that u_{k+1} appears directly, i.e.,

$$u_{k+1} - u_k\,\frac{K^*}{K^*\mathbf{1}_\Sigma}\left(\frac{f}{Ku_k}\right) + \alpha\,\frac{u_k}{K^*\mathbf{1}_\Sigma}\,p_{k+1} = 0\,, \qquad p_{k+1} \in \partial\,|u_{k+1}|_{BV(\Omega)}\,. \tag{5.22}$$

In order to verify that the iteration scheme (5.22) preserves positivity, we proceed in an analogous to the EM algorithm in Section 5.2.3. Due to the nonnegativity constraint in (5.18) the KKT conditions provide formally a Lagrange multiplier $\lambda \geq 0$, such that the stationary points of (5.19) need to fulfill

$$\begin{cases} 0 \;\in\; K^*\mathbf{1}_\Sigma - K^*\left(\dfrac{f}{Ku}\right) + \alpha\,\partial\,|u|_{BV(\Omega)} - \lambda \\ 0 \;=\; \lambda u\,. \end{cases} \tag{5.23}$$

Multiplying the first equation in (5.23) by u, the Lagrange multiplier λ can be eliminated and division by $K^*\mathbf{1}_\Sigma$ leads to a fixed point equation

$$u \;-\; u\,\frac{K^*}{K^*\mathbf{1}_\Sigma}\left(\frac{f}{Ku}\right) \;+\; \alpha\,\frac{u}{K^*\mathbf{1}_\Sigma}\,p \;=\; 0\;, \qquad p \in \partial\,|u|_{BV(\Omega)}\;, \tag{5.24}$$

which is (5.20) multiplied by u, i.e. this multiplication corresponds with the nonnegativity constraint in (5.18). Now, the iteration (5.22) is just a semi-implicit approach to (5.24). In Section 5.4.3, we prove that this iteration method actually preserves positivity if the operator K preserves positivity and the initialization u_0 is positive.

Remarkably, the second term in the iteration (5.22) is the EM step in (5.17). Consequently, the method (5.22) solving the variational problem (5.18) can be realized as a nested two-step iteration

$$\begin{cases} u_{k+\frac{1}{2}} \;=\; u_k\,\dfrac{K^*}{K^*\mathbf{1}_\Sigma}\left(\dfrac{f}{Ku_k}\right) & \text{(EM step)} \\[2ex] u_{k+1} \;=\; u_{k+\frac{1}{2}} \;-\; \alpha\,\dfrac{u_k}{K^*\mathbf{1}_\Sigma}\,p_{k+1}\;, \qquad p_{k+1} \in \partial\,|u_{k+1}|_{BV(\Omega)} & \text{(TV step)}\;. \end{cases} \tag{5.25}$$

Thus, we alternate an EM reconstruction step with a TV correction step to compute a solution of (5.18). In Section 5.3.2 we will see that this iteration scheme can be interpreted as a modified forward-backward splitting strategy. The second half step from $u_{k+\frac{1}{2}}$ to u_{k+1} can be realized by solving the convex variational problem

$$u_{k+1} \;=\; \underset{u\,\in\,BV(\Omega)}{\arg\min}\left\{\; \frac{1}{2}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u \;-\; u_{k+\frac{1}{2}})^2}{u_k} \;+\; \alpha\,|u|_{BV(\Omega)} \;\right\}\;. \tag{5.26}$$

Inspecting the first order optimality condition confirms the equivalence of this minimization with the TV correction step in (5.25). Problem (5.26) is just a modified version of the Rudin-Osher-Fatemi (ROF) model, with weight $\frac{K^*\mathbf{1}_\Sigma}{u_k}$ in the fidelity term. This analogy creates the opportunity to carry over efficient numerical schemes known for the ROF model. In the numerical realization in Chapter 5.5, we offer an algorithm analogous to Chambolle [42]. In this way, the weighted ROF problem with the exact definition of TV can be solved by using duality, obtaining an accurate, robust, and efficient algorithm.

The alternating structure of the proposed iteration (5.25) has the particular advantage that we might control the interaction between reconstruction and denoising via a simple adaption of the TV correction step. A possibility is a damped TV correction step

$$u_{k+1} \;=\; (1-\omega_k)\,u_k \;+\; \omega_k\,u_{k+\frac{1}{2}} \;-\; \omega_k\,\alpha\,\frac{u_k}{K^*\mathbf{1}_\Sigma}\,p_{k+1}\;, \qquad \omega_k \in (0,1]\;, \tag{5.27}$$

which relates the current EM iterate $u_{k+\frac{1}{2}}$ to the previous TV denoised iterate u_k in a suitable way using a damping parameter ω_k. The damped half step (5.27) can also be realized in an analogous way to (5.26), namely by minimizing the following variational problem

$$u_{k+1} \;=\; \underset{u \,\in\, BV(\Omega)}{\arg\min} \left\{ \frac{1}{2} \int_\Omega \frac{K^* \mathbf{1}_\Sigma \,(u \;-\; (\omega_k \; u_{k+\frac{1}{2}} \;+\; (1-\omega_k)\, u_k))^2}{u_k} \;+\; \omega_k \,\alpha \,|u|_{BV(\Omega)} \right\} . \tag{5.28}$$

This aspect of damping is not only motivated by numerical results, see Section 5.5, but also required to attain a monotone descent of the objective functional (5.18), respectively (5.19), see Section 5.4.4. For $\omega_k = 1$, the iteration simplifies to the original TV denoising step in (5.25). For ω_k small the iterations stay close to u_k. For an adequate choice of $\omega_k \in (0, 1]$, we can prove the convergence of the proposed two-step iteration with the damped TV denoising step, with explicit bounds on ω_k in the special case of $K = Id$ (cf. Section 5.4.4).

Further, we need appropriate stopping rules for the proposed FB-EM-TV algorithm in order to guarantee its accuracy . In addition to a maximal number of iterations, the error in the optimality condition (5.20) can be taken as a basic stopping criterion in a suitable norm. For this purpose, we define a weighted norm deduced from a weighted scalar product,

$$\langle u, v \rangle_w \;:=\; \int_\Omega u \, v \, w \; d\lambda \qquad \text{and} \qquad \|u\|_{2,w} \;:=\; \sqrt{\langle u, u \rangle_w} \;, \tag{5.29}$$

with a positive weight function w and the standard Lebesque measure λ on Ω. Hence, the error in the optimality condition can be measured reasonably in the norm

$$opt_{k+1} \;:=\; \left\| K^* \mathbf{1}_\Sigma \;-\; K^* \left(\frac{f}{K u_{k+1}} \right) \;+\; \alpha \, p_{k+1} \right\|^2_{2, u_{k+1}} . \tag{5.30}$$

Due to the fact that we use a damped two-step iteration, we are not only interested in the improvement of the whole optimality condition, but also in the convergence of the primal functions $\{u_k\}$ and the subgradients $\{p_k\}$ with $p_k \in \partial \, |u_k|_{BV(\Omega)}$. Hence, in order to establish appropriate stopping rules for these iterates, we consider the damped TV correction step (5.27) with the EM reconstruction step in (5.25),

$$u_{k+1} \;-\; \omega_k \, u_k \, \frac{K^*}{K^* \mathbf{1}_\Sigma} \left(\frac{f}{K u_k} \right) \;-\; (1 - \omega_k) \, u_k \;+\; \omega_k \, \alpha \, \frac{u_k}{K^* \mathbf{1}_\Sigma} \, p_{k+1} \;=\; 0 \; .$$

Combining this iteration scheme with the optimality condition (5.24) evaluated at u_k, which must be fulfilled in the case of convergence, we obtain an optimality statement for $\{p_k\}$ and $\{u_k\}$,

$$\alpha \; (p_{k+1} \;-\; p_k) \;+\; \frac{K^* \mathbf{1}_\Sigma \; (u_{k+1} \;-\; u_k)}{\omega_k \, u_k} \;=\; 0 \; .$$

With the aid of the weighted norm (5.29), we have now additional stopping criteria for the FB-EM-TV algorithm which guarantee the accuracy of the primal functions $\{u_k\}$ and the subgradients $\{p_k\}$,

$$u\_opt_{k+1} := \left\| \frac{K^*\mathbf{1}_\Sigma\ (u_{k+1} - u_k)}{\omega_k\ u_k} \right\|^2_{2,u_{k+1}}, \quad p\_opt_{k+1} := \| \alpha\ (p_{k+1} - p_k) \|^2_{2,u_{k+1}}. \tag{5.31}$$

We finally mention that the stopping criteria (5.30) and (5.31) are well defined, since each iterate u_k of the FB-EM-TV splitting method is strictly positive, see Lemma 5.4.11.

Based on the observations in this section we can use **Algorithm 1** to solve the problem (5.18).

Algorithm 1 FB-EM-TV

1. **Parameters:** $f,\ \alpha > 0,\ \omega \in (0, 1],\ maxEMIts > 0,\ tol > 0$
2. **Initialization:** $k = 0,\ u_0 = c > 0$

while ($k \leq maxEMIts$ **and** ($opt_k \geq tol$ **or** $u\_opt_k \geq tol$ **or** $p\_opt_k \geq tol$)) **do** ▷ (5.30), (5.31)

Compute $u_{k+\frac{1}{2}}$ via EM step in (5.25).
Set $\omega_k = \omega$.
Compute u_{k+1} via modified ROF model (5.28). ▷ Section 5.5
$k \leftarrow k + 1$

end while
return u_{k+1}

Selecting a reasonable regularization parameter α in our model is a common problem. In the case of additive Gaussian noise there exist several works in literature dealing with this problem, e.g. [117], [172], [185]. Most of them are based on the discrepancy principle and Chi-square distributions, generalized cross validation methods or unbiased predictive risk estimates. Finding an "optimal" parameter is in general more complicated for non-Gaussian noise models. Nevertheless, there exist a few works in literature addressing this issue, see e.g. [15] and the references within.

5.3.2 Forward-Backward-Splitting

In the previous section we introduced the FB-EM-TV reconstruction method as a two-step algorithm (5.25) with an additional damping modification (5.27). This two-step strategy can be interpreted surprisingly as an operator splitting algorithm.

In convex optimization such splitting methods arise in the context of decomposition problems. Recently, some works in literature picked up these splitting ideas, providing efficient algorithms in image processing, see e.g. [165], [26], [49], [48], [84]. Most of the works dealing with convex splitting strategies go back to early works of Douglas and Rachford [58], respectively of [119] and [176].

The optimality condition of our underlying variational problem (5.18), respectively (5.19), can be interpreted as a decomposition problem ($C = A + B$), regarding the convex Kullback-Leibler functional and the convex TV regularization term. Hence, we consider the stationary equation

$$0 \;\in\; C(u) \;:=\; \underbrace{K^*\mathbf{1}_\Sigma \;-\; K^*\left(\frac{f}{Ku}\right)}_{=:\,A(u)} \;+\; \underbrace{\alpha\,\partial\,|u|_{BV(\Omega)}}_{=:\,B(u)}\,, \tag{5.32}$$

with two maximal monotone operators A and B. The damped two-step iteration for the EM-TV model, (5.25) and (5.27) with $\omega_k \in (0,1]$, reads as follows

$$\begin{cases} \dfrac{K^*\mathbf{1}_\Sigma\,(\,u_{k+\frac{1}{2}} \;-\; u_k\,)}{u_k} \;+\; Au_k \;=\; 0 \\[2ex] \dfrac{K^*\mathbf{1}_\Sigma\,(\,u_{k+1} \;-\; \omega_k\,u_{k+\frac{1}{2}} \;-\; (1-\omega_k)\,u_k\,)}{u_k} \;+\; \omega_k\,B\,u_{k+1} \;=\; 0 \end{cases} \tag{5.33}$$

and can easily be reformulated as a forward-backward splitting algorithm

$$\begin{cases} \dfrac{K^*\mathbf{1}_\Sigma\,(\,\tilde{u}_{k+\frac{1}{2}} \;-\; u_k\,)}{\omega_k\,u_k} \;+\; Au_k \;=\; 0 & \text{(forward step on } A\text{)} \\[2ex] \dfrac{K^*\mathbf{1}_\Sigma\,(\,u_{k+1} \;-\; \tilde{u}_{k+\frac{1}{2}}\,)}{\omega_k\,u_k} \;+\; Bu_{k+1} \;=\; 0 & \text{(backward step on } B\text{)} \end{cases}$$

with

$$\tilde{u}_{k+\frac{1}{2}} \;=\; \omega_k\,u_{k+\frac{1}{2}} \;+\; (1-\omega_k)\,u_k\;.$$

Compared to (standard) EM-TV (5.25), in the case of damped EM-TV, the artificial time step size is therefore not only given by u_k, but can be controlled via the additional damping parameter ω_k. In a more compact form, the whole iteration can be formulated as

$$u_{k+1} \;=\; \left(I \;+\; \frac{\omega_k\,u_k}{K^*\mathbf{1}_\Sigma}\,B\right)^{-1}\left(I \;-\; \frac{\omega_k\,u_k}{K^*\mathbf{1}_\Sigma}\,A\right)u_k \;=\; (L_k \;+\; B)^{-1}\,(L_k \;-\; A)\,u_k$$

with a multiplication operator L_k defined by $\frac{K^*\mathbf{1}_\Sigma}{\omega_k\,u_k}$.

The forward-backward splitting for maximal monotone operators has been suggested independently by Lions and Mercier [119] and Passty [138]. In Section 5.4.4 we

will see that the key to proving convergence of the FB-EM-TV splitting algorithm lies in the incorporation of damping parameters. Alternatives to forward-backward splitting are the Peaceman-Rachford or Douglas-Rachford splitting schemes, see e.g. [119], which are indeed unconditionally stable, but have the numerical drawback that also an additional backward step on A has to be performed, which would mean an inversion of K in our case.

5.4 Analysis

In this chapter we carry out a mathematical analysis of the regularized problem (5.18), and prove that the problem is well-posed, that the FB-EM-TV algorithm preserves the positivity of the solution and that the proposed damped FB-EM-TV iteration scheme has a stable convergence behavior.

5.4.1 Assumptions, Definitions and Preliminary Results

At the beginning of this section, for the sake of clarity, we will repeat the properties of the operator K, having been assumed in the previous sections and we will introduce other necessary assumptions, which will be used in the following analysis.

As introduced in Section 5.2.1, K is a semi discrete operator derived from $\bar{K} : L^1(\Omega) \to L^1(\Sigma)$, which transforms, contrary to $\bar{K}$, a function from $L^1(\Omega)$ to the discrete data space $\mathbb{R}^N$. Nevertheless, to be able to present a unified theory of a continuous problem formulation (5.18), we introduced an integral formulation with point measure in Section 5.2.2. For this sake we assumed that any element g in the discrete data space $\mathbb{R}^N$ can be interpreted as samples of a function in $L^1(\Sigma)$, which we denote for the sake of convenience with g again. For the assumptions below, note that the operator $\bar{K}$ itself is linear compact with a nonclosed range and additionally preserves positivity. Hence, we assume the following:

(A1) The operator $K : L^1(\Omega) \to L^1_\mu(\Sigma)$ is linear and bounded.

(A2) The operator K satisfies $Ku \geq 0$ a.e. for any $u \geq 0$ a.e. and the equality is fulfilled if and only if $u = 0$.

(A3) The data function f is bounded and bounded away from zero, i.e. there exist positive constants c_1 and c_2 such that $0 < c_1 \leq f \leq c_2$ almost everywhere on Σ.

(A4) If $u \in L^1(\Omega)$ satisfies $c_1 \leq u \leq c_2$ a.e. for some positive constants $c_1, c_2 > 0$, then there exist $c_3, c_4 > 0$ such that $c_3 \leq Ku \leq c_4$ a.e. on Σ.

Remarks and Extensions. In the most practical situations, the assumptions (A3) and (A4) are not significantly restrictive. In the case of the "measured" data function f, the almost everywhere boundedness on Σ away from zero is reasonable, when a sufficient level of measurements has been collected. In addition, in most practical applications a certain level of background noise is available, which causes the positivity of the data.

At the first glance, the assumption (A4) is restrictive, but there are large classes of linear ill-posed problems for which the required condition is fulfilled. An example are integral equations of the first kind, which have smooth, bounded, and positive kernels. Such integral equations appear in numerous fields of applications, e.g. in geophysics and potential theory or in deconvolution problems such as fluorescence microscopy or astronomy. Another interesting example of an operator that fulfills the assumption (A4) is the X-ray transform which assigns to a function the integral values along all straight lines. This transform coincides in two dimensions with the well-known Radon transform and will be strongly applied in medical imaging. The assumption (A4) in this example is fulfilled if the length of the lines is bounded and bounded away from zero, a condition obviously satisfied in practice.

Next we give a definition to simplify the following analysis of the regularized problem (5.18).

Definition 5.4.1 (Kullback-Leibler Functional)**.** *The Kullback-Leibler (KL) functional is the function* $D_{KL} : L^1(\Sigma) \times L^1(\Sigma) \to \mathbb{R} \cup \{+\infty\}$ *with* $\Sigma \subset \mathbb{R}^m$ *bounded and measurable, given by*

$$D_{KL}(\varphi, \psi) \;=\; \int_\Sigma \varphi \log\left(\frac{\varphi}{\psi}\right) - \varphi + \psi \; d\nu \qquad \textit{for all} \qquad \varphi\,,\; \psi \;\geq\; 0 \;\textit{a.e.} \tag{5.34}$$

where ν *is a measure. Note that we use the convention* $0 \log 0 = 0$ *and that the integrand in* (5.34) *is non negative and vanishes if and only if* $\varphi = \psi$*.*

Remarks and Extensions. In the literature there exist further notations for the KL functional, like cross-entropy, information for discrimination or Kullback's I-divergence, e.g. [50], [61], [150]. The functional (5.34) generalizes the Kullback-Leibler entropy

$$D(\varphi||\psi) \;=\; \int_\Sigma \varphi \log\left(\frac{\varphi}{\psi}\right) \; d\nu$$

for functions which are not necessarily probability densities. In the definition above, the extension is realized by adding additional (linear) terms which are chosen such that (5.34) is a Bregman distance (divergence) with respect to the Boltzmann en-

tropy [150],

$$B(\psi) \;:=\; \int_\Sigma \psi \,\log \psi \;-\; \psi \;,$$
$$\text{i.e. } D_{KL}(\varphi,\psi) \;\overset{!}{=} D_B^p(\varphi,\psi) \;=\; B(\varphi) - B(\psi) - \langle p, \varphi - \psi\rangle \;\;,$$

where $p \in \partial B(\psi)$. Therefore, it shares useful properties of the Bregman distance, in particular $D_{KL}(\varphi,\psi) \geq 0$.

In the next Lemmas, we recall from [150] and [1] a collection of basic results about the KL functional and total variation $|\cdot|_{BV(\Omega)}$, which will be used in the following analysis. For further information to the both terms, we refer to [150], [61] and [1], [7], [70], [81].

Lemma 5.4.2 (Properties of KL Functional)**.** *The following statements hold:*

1. *The function $(\varphi,\psi) \mapsto D_{KL}(\varphi,\psi)$ is convex and thus also the function $(\varphi,u) \mapsto D_{KL}(\varphi,Ku)$.*
2. *For any fixed non negative $\varphi \in L^1(\Sigma)$, the function $u \mapsto D_{KL}(\varphi,Ku)$ is lower semicontinuous with respect to the weak topology of $L^1(\Sigma)$.*
3. *For any non negative function φ and ψ in $L^1(\Sigma)$, one has*

$$\|\varphi - \psi\|_1^2 \;\leq\; \left(\frac{2}{3}\|\varphi\|_1 \;+\; \frac{4}{3}\|\psi\|_1\right) D_{KL}(\varphi,\psi) \;.$$

Proof. See [150, Lemma 3.3 - 3.4]. □

Corollary 5.4.3. *If $\{\varphi_n\}$ and $\{\psi_n\}$ are bounded sequences in $L^1(\Sigma)$, then*

$$\lim_{n\to\infty} D_{KL}(\varphi_n,\psi_n) \;=\; 0 \qquad \textit{implies} \qquad \lim_{n\to\infty} \|\varphi_n \;-\; \psi_n\|_1 \;=\; 0 \;.$$

Lemma 5.4.4 (Properties of Total Variation)**.** *The following statements hold:*

1. *$|\cdot|_{BV(\Omega)}$ is convex.*
2. *$|\cdot|_{BV(\Omega)}$ is lower semicontinuous with respect to the weak topology of $L^1(\Omega)$.*
3. *Any uniformly bounded sequence $\{u_n\}$ in $BV(\Omega)$ is relatively compact in $L^1(\Omega)$.*

Proof. See [1, Theorem 2.3 - 2.5]. □

5.4.2 Well-Posedness of the Minimization Problem

In the following we verify existence, uniqueness, and stability of the minimization problem (5.18). In order to use the known properties of the KL functional from Lemma 5.4.2 for the analysis of (5.18), we add the term $f \log f \; - \; f$ in the data fidelity term. Because this expression is independent from the desired function u, the stationary points of the minimization problem are not influenced (if they exist). Thus, (5.18) is equivalent to

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \;\; a.e.}} F(u) \;:=\; D_{KL}(f, Ku) \;+\; \alpha \, |u|_{BV(\Omega)} \;, \qquad \alpha > 0 \;, \tag{5.35}$$

where D_{KL} is the Kullback-Leibler functional, see Definition 5.4.1.

For the following analysis, the compactness result from Lemma 5.4.4 is of fundamental importance. In order to use this property we introduce the following definition:

Definition 5.4.5 (Coercivity)**.** *A functional F defined on $L^1(\Omega)$ is BV-coercive (cf. [100], Definition IV.3.2.6), if the sub level sets of F are bounded in the $\|\cdot\|_{BV(\Omega)}$ norm, i.e. for all $r \in \mathbb{R}_{\geq 0}$ the set $\{u \in L^1(\Omega) \;:\; F(u) \leq r\}$ is uniform bounded in the BV norm; equivalently*

$$F(u) \;\to\; +\infty \qquad \textit{whenever} \qquad \|u\|_{BV(\Omega)} \;\to\; +\infty \;.$$

Lemma 5.4.6 (Coercivity of the Minimization Functional)**.** *Assume that the operator K does not annihilate constant functions. Then F in (5.35), is BV-coercive.*

Remarks and Extensions. According to the definition of the space of functions with bounded (total) variation, $BV(\Omega) \subset L^1(\Omega)$ is valid and we can extend the admissible solution set of the minimization problem (5.35) from $BV(\Omega)$ to $L^1(\Omega)$. For this sake, we extend the total variation to a functional on $L^1(\Omega)$ by $|u|_{BV(\Omega)} = \infty$ for $u \in L^1(\Omega) \setminus BV(\Omega)$, where furthermore solutions from $BV(\Omega)$ are preferred during minimization.

Because K is linear, see (A1), the condition in Lemma 5.4.6 is equivalent to

$$K \mathbf{1}_\Omega \;\neq\; 0 \tag{5.36}$$

where $\mathbf{1}_\Omega$ is the characteristic function on Ω.

Proof of Lemma 5.4.6. For the proof we derive an estimate of the form

$$\|u\|_{BV(\Omega)} \;=\; ||u||_1 \;+\; |u|_{BV(\Omega)} \;\leq\; c_1 \, [F(u)]^2 \;+\; c_2 \, F(u) \;+\; c_3 \tag{5.37}$$

with constants $c_1 \geq 0$, $c_2 > 0$ and $c_3 \geq 0$. Then, the desired coercivity property follows directly from the positivity of the functional F for all $u \in L^1(\Omega)$ with $u \geq 0$ *a.e.*

For the derivation of this estimate we use that any $u \in BV(\Omega)$ has a decomposition of the form

$$u \;=\; w \,+\, v \;, \tag{5.38}$$

where

$$w \;=\; \left(\frac{\int_\Omega u \; dx}{|\Omega|}\right) \mathbf{1}_\Omega \qquad \text{and} \qquad v \;:=\; u \,-\, w \qquad \text{with} \qquad \int_\Omega v \; dx \;=\; 0 \;. \tag{5.39}$$

First, we estimate $|v|_{BV(\Omega)}$ and $||v||_1$. Because the constant functions have no variation, the positivity of the KL functional yields

$$\alpha\, |v|_{BV(\Omega)} \;\leq\; \alpha\, |u|_{BV(\Omega)} \;\leq\; F(u) \qquad \text{which implies} \qquad |v|_{BV(\Omega)} \;\leq\; \frac{1}{\alpha}\, F(u) \;.$$

Together with the Poincaré-Wirtinger inequality, see e.g. [7], this yields an estimate of the L^1 norm,

$$||v||_1 \;\leq\; C_1 |v|_{BV(\Omega)} \;\leq\; C_1 \,\frac{1}{\alpha}\, F(u) \;, \tag{5.40}$$

where $C_1 > 0$ is a constant that depends from $\Omega \subset \mathbb{R}^n$ and n only. If we use the decomposition (5.38) and the results to $|v|_{BV(\Omega)}$ and $||v||_1$, then the problem (5.37) reduces to the estimation of the L^1 norm of constant functions,

$$\begin{aligned} \|u\|_{BV(\Omega)} \;&\leq\; ||w||_1 \,+\, ||v||_1 \,+\, |v|_{BV(\Omega)} \\ &\leq\; ||w||_1 \,+\, (C_1 + 1)\, \frac{1}{\alpha}\, F(u) \;. \end{aligned} \tag{5.41}$$

For this purpose, we consider the L^1 distance between $Ku = Kw + Kv$ and f. With Lemma 5.4.2 (3) we obtain an estimate from above,

$$\begin{aligned} ||(Kv - f) \,+\, Kw||_1^2 \;&\leq\; \left(\frac{2}{3}||f||_1 \,+\, \frac{4}{3}||Kv + Kw||_1\right) D_{KL}(f, Ku) \\ &\leq\; \left(\frac{2}{3}||f||_1 \,+\, \frac{4}{3}||Kv||_1 \,+\, \frac{4}{3}||Kw||_1\right) F(u) \;, \end{aligned}$$

as an estimate from below we obtain

$$\begin{aligned} ||(Kv - f) \,+\, Kw||_1^2 \;&\geq\; (||Kv - f||_1 \,-\, ||Kw||_1)^2 \\ &\geq\; ||Kw||_1 \,(||Kw||_1 \,-\, 2\,||Kv - f||_1) \;. \end{aligned}$$

Combining (5.40) with both inequalities yields

$$||Kw||_1 \left(||Kw||_1 - 2 \left(||K|| C_1 \frac{1}{\alpha} F(u) + ||f||_1 \right) \right) \leq \left(\frac{2}{3}||f||_1 + \frac{4}{3}||K|| C_1 \frac{1}{\alpha} F(u) + \frac{4}{3}||Kw||_1 \right) F(u) \ . \quad (5.42)$$

This expression contains up to now terms which describe the function w in dependence of the operator K only. For the estimate of $||w||_1$, we use the assumption (5.36) on the operator K. Thus, there exists a constant $C_2 > 0$ with

$$C_2 \ = \ \frac{\int_\Sigma |K\mathbf{1}_\Omega| \, dx}{|\Omega|} \qquad \text{and} \qquad ||Kw||_1 \ = \ C_2 ||w||_1 \ .$$

This identity used in the inequality (5.42) yields

$$C_2||w||_1 \left(C_2||w||_1 - 2 \left(||K|| C_1 \frac{1}{\alpha} F(u) + ||f||_1 \right) - \frac{4}{3} F(u) \right) \leq \left(\frac{2}{3}||f||_1 + \frac{4}{3}||K|| C_1 \frac{1}{\alpha} F(u) \right) F(u) \ . \quad (5.43)$$

To obtain an estimate of the form (5.37), we distinguish between two cases:

Case 1: If

$$C_2 \, ||w||_1 \ - \ 2 \left(||K|| \, C_1 \, \frac{1}{\alpha} \, F(u) \ + \ ||f||_1 \right) \ - \ \frac{4}{3} F(u) \ \geq \ 1 \ , \quad (5.44)$$

then from (5.43) we conclude

$$||w||_1 \ \leq \ \frac{1}{C_2} \left(\frac{2}{3} \, ||f||_1 \ + \ \frac{4}{3} \, ||K|| \, C_1 \, \frac{1}{\alpha} \, F(u) \right) F(u)$$

and we obtain with (5.41)

$$\|u\|_{BV(\Omega)} \ \leq \ \frac{4 \, C_1 \, ||K||}{3 \, C_2 \, \alpha} \, [F(u)]^2 \ + \ \left(\frac{2}{3 \, C_2} \, ||f||_1 \ + \ \frac{C_1 + 1}{\alpha} \right) F(u) \ . \quad (5.45)$$

Case 2: If the condition (5.44) does not hold, i.e.

$$||w||_1 \ < \ \frac{1}{C_2} \left(1 \ + \ 2 \left(||K|| \, C_1 \, \frac{1}{\alpha} \, F(u) \ + \ ||f||_1 \right) \ + \ \frac{4}{3} \, F(u) \right) \ ,$$

then from (5.41) we find

$$\|u\|_{BV(\Omega)} \ \leq \ \left(\frac{2 \, ||K|| \, C_1 \, \frac{1}{\alpha} + \frac{4}{3}}{C_2} \ + \ \frac{C_1 + 1}{\alpha} \right) F(u) \ + \ \frac{1 + 2 \, ||f||_1}{C_2} \ . \quad (5.46)$$

With the assumptions (A1) and (A3) is f in $L^1(\Sigma)$ and $||K|| < \infty$ and we obtain from (5.45) and (5.46) the desired coercivity property. □

Theorem 5.4.7 (Existence of Minimizers)**.** *Let the operator K satisfy the assumptions of Lemma 5.4.6. Then, the functional F defined in (5.35) has a minimizer in $BV(\Omega)$.*

Proof. For the proof, we use the direct method of the calculus of variations, see e.g. [7, Section 2.1.2]: Let $\{u_n\} \subset BV(\Omega)$, $u_n \geq 0$ *a.e.*, be a minimizing sequence for the functional F, i.e.

$$\lim_{n\to\infty} F(u_n) \;=\; \inf_{u \in BV(\Omega)} F(u) \;=:\; F_{\min} \;<\; \infty \,.$$

With the assumptions on the operator K, Lemma 5.4.6 implies that the functional F is BV-coercive. Hence, all elements u_n of the sequence are uniformly bounded in the BV norm. As a consequence of the compactness result from Lemma 5.4.4 (3), there exists a subsequence $\{u_{n_j}\}$ which converges to some $\tilde{u} \in L^1(\Omega)$. Actually, the function $\tilde{u}$ lies in $BV(\Omega)$, since $|\cdot|_{BV(\Omega)}$ is lower semicontinuous (see Lemma 5.4.4 (2)) and the sequence $\{u_n\}$ is uniformly bounded in $BV(\Omega)$.

Moreover, from Lemma 5.4.2 (2) and 5.4.4 (2) we know that the functional F is lower semicontinuous, i.e.

$$F(\tilde{u}) \;\leq\; \liminf_{j\to\infty} F(u_{n_j}) \;=\; F_{\min} \,.$$

Obviously, this inequality implies that $\tilde{u}$ is a minimizer of F. □

Next, we consider the uniqueness of the minimizers, for which it suffices to verify the strict convexity of the objective functional. It is straight-forward to see that the negative log is strictly convex and consequently the Kullback-Leibler divergence is strictly convex with respect to u if K is injective, i.e. the null space is $\mathcal{N}(K) = \{0\}$. Since the regularization term is assumed convex (see Lemma 5.4.4) we can immediately conclude the following result:

Theorem 5.4.8 (Uniqueness of Minimizers)**.** *Let K be an injective operator and $f > 0$. Then, the function $u \mapsto D_{KL}(f, Ku)$ and also the functional F from (5.35) are strictly convex. In particular the minimizer of F is unique in $BV(\Omega)$.*

After existence and uniqueness of minimizers we show below the stability of the regularized problem (5.35) with respect to a certain kind of data perturbations. In Subsection 5.2.1 we already described that the given measurements are typically discrete and can be interpreted in our framework as averages of a function $f \in L^1(\Sigma)$. The open question is certainly the suitable choice of the function f, e.g. the interpolation type of the measurements. Moreover, the physically limited discrete

construction of the detectors leads to a natural loss of information because not all signals can be acquired. Consequently, a stability result is required that guarantees that the regularized approximations converge to the solution u if e.g. the interpolated measurements converge to a preferably smooth function f. Because the measurements are still a realization of Poisson distributed random variables, it is natural to assess the convergence in terms of the KL functional, as detailed below in equality (5.47).

Remarks and Extensions. The assumption (A3) guarantees not only that the data function f has a finite L^∞ norm, but also that $\log f$ belongs to $L^\infty(\Sigma)$. This fact will be needed in the subsequent Theorem.

Theorem 5.4.9 (Stability with respect to perturbations in the data)**.** *Let $\alpha > 0$ be fixed and suppose that $f_n \in L^1(\Sigma)$, $n \in \mathbb{N}$, are nonnegative approximations of a data function f with*

$$\lim_{n\to\infty} D_{KL}(f_n, f) \;=\; 0 \;. \tag{5.47}$$

Moreover, let

$$u_n \;\in\; \operatorname*{arg\,min}_{\substack{u \in BV(\Omega) \\ u \geq 0 \;\; a.e.}} F_n(u) \;:=\; D_{KL}(f_n, Ku) \;+\; \alpha\, |u|_{BV(\Omega)} \;, \qquad n \in \mathbb{N} \;, \tag{5.48}$$

and u is a solution of the regularized problem (5.35) corresponding to the data function f. In addition, we assume that the operator K does not annihilate constant functions and that $\log Ku$ belongs to the function space $L^\infty(\Sigma)$, i.e. there exist positive constants c_1 and c_2 such that

$$0 \;<\; c_1 \;\leq\; Ku \;\leq\; c_2 \qquad \text{almost everywhere on } \Sigma \;. \tag{5.49}$$

Then, the problem (5.35) is stable with respect to the perturbations in the data, i.e. the sequence $\{u_n\}$ has a convergent subsequence and every convergent subsequence converges in the L^1 norm to a minimizer of the functional F in (5.35).

Proof. For the existence of a convergent subsequence we want to use the compactness result from Lemma 5.4.4. To this end, we show that the sequence $\{F_n(u_n)\}$ is uniformly bounded and that the functionals F_n are uniformly BV-coercive.

Firstly, we show the uniform boundedness of the sequence $\{F_n(u_n)\}$. Let $\alpha > 0$ be a fixed regularization parameter. For any $n \in \mathbb{N}$, the definition of u_n implies

$$F_n(u_n) \;=\; D_{KL}(f_n, Ku_n) \;+\; \alpha\, |u_n|_{BV(\Omega)} \;\leq\; D_{KL}(f_n, Ku) \;+\; \alpha\, |u|_{BV(\Omega)} \;. \tag{5.50}$$

Hence, the sequence $\{F_n(u_n)\}$ is bounded if the sequence $\{D_{KL}(f_n, Ku)\}$ on the right-hand side of (5.50) is bounded. To this end, we use the condition (5.47) and

Corollary 5.4.3, so the sequence $\{f_n\}$ converges strongly to f in $L^1(\Sigma)$ as well as pointwise almost everywhere. Thus, the assumptions (5.49) and (A3) together with the inequality

$$\begin{aligned} |D_{KL}(f_n, Ku) \;-\; D_{KL}(f, Ku) \;-\; D_{KL}(f_n, f)| \;&=\; \left| \int_\Sigma (\log Ku - \log f)(f - f_n)\, d\mu \right| \\ &\leq\; ||\log Ku \;-\; \log f||_\infty ||f \;-\; f_n||_1 \end{aligned}$$

imply the convergence

$$\lim_{n\to\infty} D_{KL}(f_n, Ku) \;=\; D_{KL}(f, Ku)\;. \tag{5.51}$$

Because u is a minimizer of the regularized problem (5.35) corresponding to the data function f, the expressions $D_{KL}(f, Ku)$ and $|u|_{BV(\Omega)}$ are bounded and therefore also the sequence $\{D_{KL}(f_n, Ku)\}$ is bounded, since it converges. This fact together with (5.50), yields the uniform boundedness of the sequence $\{F_n(u_n)\}$.

Next, we prove that the regularized functionals F_n are uniformly BV-coercive, i.e. for any sequence $\{u_n\}$ in $L^1(\Omega)$ with $u_n \geq 0$ *a.e.*,

$$F_n(u_n) \;\to\; +\infty \qquad \text{whenever} \qquad \|u_n\|_{BV(\Omega)} \;\to\; +\infty\;.$$

For the proof we put $u_n = w_n + v_n$ as in equation (5.38) and (5.39), and repeat the proof of Lemma 5.4.6 with u_n and F_n instead of u and F. Since the operator K does not annihilate constant functions, we obtain $||Kw_n||_1 \;=\; C_2||w_n||_1$ with C_2 as in the proof of Lemma 5.4.6 and as in (5.43),

$$\begin{aligned} C_2\,||w_n||_1 \left(C_2\,||w_n||_1 \;-\; 2\,\left(||K||\,C_1\,\frac{1}{\alpha}\,F_n(u_n) \;+\; ||f_n||_1 \right) \;-\; \frac{4}{3}\,F_n(u_n) \right) \\ \leq\; \left(\frac{2}{3}\,||f_n||_1 \;+\; \frac{4}{3}\,||K||\,C_1\,\frac{1}{\alpha}\,F_n(u_n) \right) F_n(u_n)\;. \end{aligned}$$

Since the sequence $\{f_n\}$ converges strongly to f in $L^1(\Sigma)$, it is also bounded in the L^1 norm. The upper bound on each $||f_n||_1$ and the boundedness of the operator norm of K yield uniform BV coercivity as in the two cases in the proof of Lemma 5.4.6.

The uniform BV-coercivity together with the boundedness of the sequence $\{F_n(u_n)\}$ implies that the sequence $\{u_n\}$ is uniformly bounded in the BV norm. Then, Lemma 5.4.4 (3) ensures the existence of a subsequence $\{u_{n_j}\}$ converging strongly to some $\tilde{u} \in L^1(\Omega)$. Actually, the function $\tilde{u}$ lies in $BV(\Omega)$, since $|\cdot|_{BV(\Omega)}$ is lower semicontinuous with respect to the weak topology of $L^1(\Omega)$ (see Lemma 5.4.4 (2)), i.e.

$$|\tilde{u}|_{BV(\Omega)} \;\leq\; \liminf_{j\to\infty} |u_{n_j}|_{BV(\Omega)} \;<\; \infty\;.$$

Now let $\{u_{n_j}\}$ denote an arbitrary subsequence of $\{u_n\}$ that converges strongly to $\tilde{u} \in L^1(\Omega)$. The boundedness of the operator K (see (A1)) implies the strong convergence of the sequence $\{Ku_{n_j}\}$ to $K\tilde{u}$ in $L^1(\Sigma)$ and hence, pointwise convergence almost everywhere. Since all f_n and u_n are nonnegative and K is an operator that preserves positivity (see (A2)), Fatou's Lemma can be applied to the sequence $\{f_{n_j} \log(f_{n_j}/Ku_{n_j}) - f_{n_j} + Ku_{n_j}\}$ and yields

$$D_{KL}(f, K\tilde{u}) \leq \liminf_{j\to\infty} D_{KL}(f_{n_j}, Ku_{n_j}) \,. \tag{5.52}$$

Due to the lower semicontinuity of the functional $|\cdot|_{BV(\Omega)}$ (see Lemma 5.4.4 (2)) and due to (5.50), (5.51) and (5.52), we obtain

$$\begin{aligned}
D_{KL}(f, K\tilde{u}) + \alpha\,|\tilde{u}|_{BV(\Omega)} &\overset{(5.52)}{\leq} \liminf_{j\to\infty} D_{KL}(f_{n_j}, Ku_{n_j}) + \alpha \liminf_{j\to\infty} |u_{n_j}|_{BV(\Omega)} \\
&\leq \liminf_{j\to\infty} \left[D_{KL}(f_{n_j}, Ku_{n_j}) + \alpha\,|u_{n_j}|_{BV(\Omega)} \right] \\
&\leq \limsup_{j\to\infty} \left[D_{KL}(f_{n_j}, Ku_{n_j}) + \alpha\,|u_{n_j}|_{BV(\Omega)} \right] \\
&\overset{(5.50)}{\leq} \limsup_{j\to\infty} \left[D_{KL}(f_{n_j}, Ku) + \alpha\,|u|_{BV(\Omega)} \right] \\
&\overset{(5.51)}{=} D_{KL}(f, Ku) + \alpha\,|u|_{BV(\Omega)} \,.
\end{aligned}$$

This means that $\tilde{u}$ is a minimizer of problem (5.35). □

Remarks and Extensions. For the proof of stability, the condition (5.49) is required, which means that $\log Ku$ lies in $L^\infty(\Sigma)$ where u is a regularized solution of the minimization problem (5.18). Due to the assumption (A4), it suffices to claim that u is bounded and bounded away from zero. The authors in [150] prove that this condition is satisfied, if the total variation in (5.35) is replaced by the KL functional $D_{KL}(\cdot, u^*)$ as the regularization, where u^* denotes a prior estimate of the solution satisfying the same boundedness condition. In the case of total variation regularization it is more difficult to prove a similar property.

In Section 5.4.3, we can show at least that the sequence $\{u_k\}$ of the iterates of the FB-EM-TV splitting algorithm (5.25) has the boundedness property and the boundedness away from zero, assumed that condition (A3) on the data is fulfilled and the initialization function u_0 is strictly positive. For this reason, we think that (5.49) is a reasonable assumption.

Analogous to the reasoning above, the convergence of the subsequences in Theorem 5.4.9 can also be proved in the L^p norm with $1 \leq p < d/(d-1)$, since any uniformly bounded sequence $\{u_n\}$ in $BV(\Omega)$ is actually relatively compact in $L^p(\Omega)$ for $1 \leq p < d/(d-1)$, see [1, Theorem 2.5].

As in Theorem 5.4.9 it is also possible to consider perturbations of the operator K. The proof is similar and only slight modifications are necessary. However, different assumptions to the perturbed operators K_n are needed, see e.g. in [1, Theorem 4.2]. Unfortunately it is also essential that the operators K_n fulfill the assumption (5.49), i.e. $K_n u$ is bounded and bounded away from zero for any $n \in \mathbb{N}$ where u is a regularized solution of the minimization problem (5.18), which severely restricts the possible perturbations of the operator K.

We finally mention that stability estimates for this problem have been derived in [16] in a different setting. There the assumptions on the possible data perturbations are more restrictive (in the supremum norm), while the assumptions on the operator perturbations are relaxed.

5.4.3 Positivity Preservation of FB-EM-TV

In the following we further discuss the properties of the iterates u_k during the FB-EM-TV algorithm. Given a positive $u_k \in BV(\Omega)$ it is straight-forward to see that $u_{k+\frac{1}{2}}$ is well defined and nonnegative due to the properties of the EM algorithm and the assumptions (A3) and (A4). An existence and uniqueness proof for the second half step, analogous to the classical results for the ROF model, yields also the existence of $u_{k+1} \in BV(\Omega)$. In order to show inductively the well-definedness of the iterates it remains to verify that u_{k+1} is indeed positive. Note that if any u_k is negative during the iteration, the objective functional in the second half step of the iteration is not convex anymore and hence the existence and uniqueness of u_{k+1} cannot be guaranteed. The non-negativity of a solution is also desired in our reconstruction models, since in typical applications the functions represent densities or intensity information. This aspect is considered explicitly by using the constrained optimization problem (5.18). To clarify the positivity preservation we present a maximum principle for the weighted ROF problem, i.e. for the second half step of the forward-backward splitting strategy:

Lemma 5.4.10 (Maximum Principle for Weighted ROF Denoising)**.** *Let $\tilde{u} \in BV(\Omega)$ be the minimizer of the variational problem*

$$\min_{u \in BV(\Omega)} J(u) := \frac{1}{2} \int_\Omega \frac{(u - q)^2}{h} + \beta |u|_{BV(\Omega)} , \qquad \beta > 0 , \tag{5.53}$$

where $q \in L^\infty(\Omega)$ and the weighting function h are positive. Then the following maximum principle holds

$$0 < \inf q \leq \inf \tilde{u} \leq \sup \tilde{u} \leq \sup q . \tag{5.54}$$

Proof. Let $\tilde{u}$ be a minimizer of J. For the proof we show that it exists a function v with

$$0 \;<\; \inf q \;\leq\; \inf v \;\leq\; \sup v \;\leq\; \sup q \tag{5.55}$$

and

$$J(v) \;\leq\; J(\tilde{u}) \;. \tag{5.56}$$

Then, the desired property (5.54) follows directly from the strict convexity of the functional in (5.53), i.e from the uniqueness of the solution.

We define the function v as a version $\tilde{u}$ cut off at $\inf q$ and $\sup q$, i.e.

$$v \;:=\; \min\left\{\max\left\{\tilde{u}, \inf q\right\}, \sup q\right\} \;.$$

With this definition, the property (5.55) is directly guaranteed. To show (5.56), we use

$$M \;:=\; \left\{\, x \;\in\; \Omega \mid v(x) \;=\; \tilde{u}(x) \,\right\} \;\subseteq\; \Omega$$

and estimate first the total variation of v by the total variation of $\tilde{u}$. The function v has (due to its definition) no variation on $\Omega \setminus M$ and we obtain

$$|v|_{BV(\Omega)} \;=\; |v|_{BV(M)} \;=\; |\tilde{u}|_{BV(M)} \;\leq\; |\tilde{u}|_{BV(\Omega)} \;. \tag{5.57}$$

The corresponding data fidelity terms can be estimated as follows: First, due to definition of v the data fidelity terms agree on M. In case of $x \in \Omega \setminus M$, we distinguish between two cases:

Case 1: If $\tilde{u}(x) \;\geq\; \sup q$, then $v(x) \;=\; \sup q$ and

$$0 \;\leq\; v(x) \;-\; q(x) \;=\; \sup q \;-\; q(x) \;\leq\; \tilde{u}(x) \;-\; q(x) \;,$$

which implies $\quad (v(x) \;-\; q(x))^2 \;\leq\; (\tilde{u}(x) \;-\; q(x))^2 \quad .$

Case 2: If $\tilde{u}(x) \;\leq\; \inf q$, then $v(x) \;=\; \inf q$ and

$$0 \;\leq\; -v(x) \;+\; q(x) \;=\; -\inf q \;+\; q(x) \;\leq\; -\tilde{u}(x) \;+\; q(x) \;,$$

which implies $\quad (v(x) \;-\; q(x))^2 \;\leq\; (\tilde{u}(x) \;-\; q(x))^2 \quad .$
Finally, we obtain

$$(v \;-\; q)^2 \;\leq\; (\tilde{u} \;-\; q)^2 \;, \qquad \forall x \;\in\; \Omega \;.$$

Now, the property (5.56) is fulfilled due to the positivity of the weighting function h and (5.57). □

Lemma 5.4.11 (Positivity of the FB-EM-TV Algorithm)**.** *Let $u_0 > 0$ and let the assumptions (A3) and (A4) be fulfilled. Then each iterate of the FB-EM-TV splitting algorithm in (5.25), i.e. $u_{k+\frac{1}{2}}$ and u_{k+1} for $k = 0, 1, \dots$, is strictly positive.*

Proof. Since $u_0 > 0$, $f > 0$ and the operator K and therewith the adjoint operator K^* does not affect the strict positivity, the first EM reconstruction step $u_{\frac{1}{2}}$ in (5.25) is strictly positive. Because the TV correction step in (5.25) can be realized via the weighted ROF problem (5.26), the maximum principle in Lemma 5.4.10 using $q := u_{\frac{1}{2}} > 0$ and $h := \frac{u_0}{K^*\mathbf{1}_\Sigma} > 0$ yields $u_1 > 0$. Inductively, the strict positivity of the subsequent iterates $u_{k+\frac{1}{2}}$ and u_{k+1} for $k = 1, 2, \dots$ is obtained by the same arguments. □

5.4.4 Convergence Results

In Section 5.3.2 we provided an interpretation of the (damped) FB-EM-TV reconstruction method in terms of a forward-backward operator splitting algorithm. In the past, several works in convex analysis dealt with the convergence of splitting strategies for solving decomposition problems, see e.g. Tseng [176] and Gabay [76]. For the proposed algorithm

$$u_{k+1} \;=\; \left(I \;+\; \frac{\omega_k\, u_k}{K^*\mathbf{1}_\Sigma}\, B\right)^{-1} \left(I \;-\; \frac{\omega_k\, u_k}{K^*\mathbf{1}_\Sigma}\, A\right)\, u_k \;,$$

Gabay provided a proof of weak convergence of the forward-backward splitting approach under the assumption of a fixed damping parameter strictly less than twice the modulus of A^{-1}. On the other hand, Tseng later gave a convergence proof, where applied to our case, the damping values $\frac{\omega_k u_k}{K^*\mathbf{1}_\Sigma}$ need to be bounded in the following way:

$$\epsilon \;\leq\; \frac{\omega_k\, u_k}{K^*\mathbf{1}_\Sigma} \;\leq\; 4m \;-\; \epsilon \;, \qquad \epsilon \;\in\; (0, 2m] \;,$$

where the first functional needs to be strictly convex with modulus m. These assumptions cannot be verified in our case, due to the modulus assumption on the data fidelity and due to the upper bounds for u_k.

For these reasons we searched for another criterion based on the damping strategy, to guarantee a monotone descent of the objective functional (5.18) in the FB-EM-TV algorithm.

In the following theorem we will establish the convergence of the damped FB-EM-TV splitting algorithm under appropriate assumptions on the damping parameters ω_k.

Theorem 5.4.12 (Convergence of the damped FB-EM-TV algorithm)**.** *Let K be an injective operator and let $\{u_k\}$ be the sequence of iterates obtained from the damped FB-EM-TV algorithm (5.33). If there exists a sequence of corresponding damping parameters $\{\omega_k\}$, $\omega_k \in (0,1]$, satisfying the inequality*

$$\omega_k \;\leq\; \frac{\displaystyle\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}\,-\,u_k)^2}{u_k}\,d\lambda}{\displaystyle\sup_{v\,\in\,[u_k,u_{k+1}]}\frac{1}{2}\int_\Sigma \frac{f\,(Ku_{k+1}\,-\,Ku_k)^2}{(Kv)^2}\,d\mu}\;(1\,-\,\epsilon)\;,\qquad \epsilon\;\in\;(0,1)\;, \tag{5.58}$$

then the objective functional F defined in (5.35) is decreasing during the iteration. If in addition, the function $K^*\mathbf{1}_\Sigma$, the damping parameters and the iterates are bounded away from zero by positive constants c_1, c_2 and c_3 such that

$$0\;<\;c_1\;\leq\;K^*\mathbf{1}_\Sigma\;,\qquad 0\;<\;c_2\;\leq\;\omega_k\qquad\text{and}\qquad 0\;<\;c_3\;\leq\;u_k\;,\qquad \forall\,k\geq 0\;, \tag{5.59}$$

then the iteration method converges to a minimizer of the functional F in the weak topology on $BV(\Omega)$ and in the strong topology on $L^1(\Omega)$.*

Proof.
First step: Monotone descent of the objective functional
To get a descent of the objective functional F using an adequate damping strategy, we look for a condition on the damping parameters $\{\omega_k\}$, which guarantees

$$F(u_{k+1})\;+\;\frac{\epsilon}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}\,-\,u_k)^2}{u_k}\,d\lambda\;\leq\;F(u_k)\;,\qquad \forall\,k\geq 0\;. \tag{5.60}$$

For this purpose, we start with the TV denoising step in (5.33), multiply it with $u_{k+1}-u_k$ and integrate over Ω. Thus, for $p_{k+1}\in\partial\,|u_{k+1}|_{BV(\Omega)}$, we obtain

$$\begin{aligned} 0\;&=\;\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}\;-\;\omega_k\,u_{k+\frac{1}{2}}\;-\;(1-\omega_k)\,u_k)\,(u_{k+1}-u_k)}{u_k}\,d\lambda \\ &\qquad\qquad +\;\omega_k\,\alpha\,\langle p_{k+1},u_{k+1}-u_k\rangle \\ &=\;\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\;+\;\omega_k\,\frac{K^*\mathbf{1}_\Sigma\,(u_k-u_{k+\frac{1}{2}})\,(u_{k+1}-u_k)}{u_k}\,d\lambda \\ &\qquad\qquad +\;\omega_k\,\alpha\,\langle p_{k+1},u_{k+1}-u_k\rangle\;. \end{aligned}$$

Due to the definition of subgradients in (3.4), we have

$$\langle p_{k+1},u_{k+1}-u_k\rangle\;\geq\;|u_{k+1}|_{BV(\Omega)}-|u_k|_{BV(\Omega)}$$

and

$$\alpha\,|u_{k+1}|_{BV(\Omega)} \;-\; \alpha\,|u_k|_{BV(\Omega)} \;+\; \frac{1}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\,d\lambda \\ \leq\; -\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_k-u_{k+\frac{1}{2}})\,(u_{k+1}-u_k)}{u_k}\,d\lambda\;.$$

Adding the difference $D_{KL}(f,Ku_{k+1})-D_{KL}(f,Ku_k)$ on both sides of this inequality and considering the definition of the KL functional D_{KL} in (5.34) and the objective functional F in (5.35) yields

$$\begin{aligned} &F(u_{k+1}) \;-\; F(u_k) \;+\; \frac{1}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\,d\lambda \\ &\leq \int_\Sigma \left[f\log\left(\frac{f}{Ku_{k+1}}\right) \;+\; Ku_{k+1} \;-\; f\log\left(\frac{f}{Ku_k}\right) \;-\; Ku_k\right]\,d\mu \\ &\qquad -\int_\Omega \left[K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k) \;-\; \frac{K^*\mathbf{1}_\Sigma\,u_{k+\frac{1}{2}}}{u_k}\,(u_{k+1}-u_k)\right]\,d\lambda \qquad (5.61)\\ &= \int_\Sigma \left[f\log\left(\frac{f}{Ku_{k+1}}\right) \;-\; f\log\left(\frac{f}{Ku_k}\right)\right]\,d\mu \;-\; \int_\Omega \left[K^*\left(\frac{f}{Ku_k}\right)(u_{k+1}-u_k)\right]\,d\lambda\;. \end{aligned}$$

The last equality in (5.61) holds since

$$\int_\Sigma Ku\,d\mu \;=\; \langle\,Ku,\mathbf{1}_\Sigma\,\rangle_\Sigma \;=\; \langle\,u,K^*\mathbf{1}_\Sigma\,\rangle_\Omega \;=\; \int_\Omega K^*\mathbf{1}_\Sigma\,u\,d\lambda$$

and $u_{k+\frac{1}{2}}$ is given by the EM reconstruction step in (5.25). Defining the functional G as

$$G(u) \;:=\; \int_\Sigma f\log\left(\frac{f}{Ku}\right)\,d\mu$$

and $\phi_1(t) \;:=\; G(u+t\,w_1)$ for any $w_1\in L^1(\Omega)$, the Gáteaux derivative dG of G in direction w_1 is given by

$$\begin{aligned} dG(u;\,w_1) \;=\; \phi_1'(t)\,|_{t=0} \;&=\; \int_\Sigma \frac{d}{dt}\left(f\log\left(\frac{f}{Ku+t\,Kw_1}\right)\right)\,d\mu\,\Big|_{t=0} \\ &=\; \left\langle -\frac{f}{Ku},Kw_1\right\rangle_\Sigma \;=\; \left\langle -K^*\left(\frac{f}{Ku}\right),w_1\right\rangle_\Omega\;. \end{aligned}$$

Interpreting the right-hand side of inequality (5.61) as a Taylor linearization of G yields

$$\begin{aligned} &F(u_{k+1}) \;-\; F(u_k) \;+\; \frac{1}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\,d\lambda \\ &\qquad\leq\; G(u_{k+1}) \;-\; G(u_k) \;-\; dG(u_k;\,u_{k+1}-u_k) \\ &\qquad=\; \frac{1}{2}\,d^2G(v;\,u_{k+1}-u_k,\,u_{k+1}-u_k)\;,\qquad v\;\in\;[u_k,u_{k+1}]\;, \qquad (5.62)\\ &\qquad\leq\; \sup_{v\,\in\,[u_k,u_{k+1}]}\;\frac{1}{2}\,d^2G(v;\,u_{k+1}-u_k,\,u_{k+1}-u_k)\;. \end{aligned}$$

We can compute the second Gáteaux derivative of G,

$$d^2G(u;\, w_1,\, w_2) \;=\; -\int_\Sigma \frac{d}{dt}\left(\frac{f}{Ku + t\,Kw_2}\,Kw_1\right)\,d\mu\bigg|_{t=0} \;=\; \int_\Sigma \frac{f\,Kw_2\,Kw_1}{(Ku)^2}\,d\mu\;.$$

Plugging the computed derivative in the inequality (5.62), we obtain

$$F(u_{k+1}) \;-\; F(u_k) \;+\; \frac{1}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\,d\lambda \\ \leq \sup_{v\,\in\,[u_k,u_{k+1}]} \frac{1}{2}\int_\Sigma \frac{f\,(Ku_{k+1} \,-\, Ku_k)^2}{(Kv)^2}\,d\mu\;. \quad (5.63)$$

Finally, we split the third term on the left-hand side of (5.63) with $\epsilon \in (0,1)$,

$$F(u_{k+1}) \;+\; \frac{\epsilon}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\,d\lambda \;+\; \frac{1-\epsilon}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\,d\lambda \\ \leq \sup_{v\,\in\,[u_k,u_{k+1}]} \frac{1}{2}\int_\Sigma \frac{f\,(Ku_{k+1} \,-\, Ku_k)^2}{(Kv)^2}\,d\mu \;+\; F(u_k)\;,$$

and obtain the condition (5.60), i.e. a descent of the objective functional F, if

$$\sup_{v\,\in\,[u_k,u_{k+1}]} \frac{1}{2}\int_\Sigma \frac{f\,(Ku_{k+1} \,-\, Ku_k)^2}{(Kv)^2}\,d\mu \;\leq\; \frac{1-\epsilon}{\omega_k}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1}-u_k)^2}{u_k}\,d\lambda\;. \quad (5.64)$$

By solving (5.64) for w_k, we obtain the condition (5.58) for the damping parameters $\{\omega_k\}$. By a suitable choice of ϵ in (5.58), we can additionally guarantee $\omega_k \;\leq\; 1$, $\forall k \;\geq\; 0$.

Second step: Convergence of the primal iterates

Next, in order to show that the iteration method converges to a minimizer of the functional F, we need a convergent subsequence of the primal iterates $\{u_k\}$. Since the operator K is injective, the functional F is BV-coercive according to Lemma 5.4.6 and we obtain from (5.37),

$$\|u_k\|_{BV(\Omega)} \;\leq\; c_4\,[F(u_k)]^2 \;+\; c_5\,F(u_k) \;+\; c_6 \;\leq\; c_4\,[F(u_0)]^2 \;+\; c_5\,F(u_0) \;+\; c_6\;,$$

for all $k \;\geq\; 0$, and with constants $c_4 \geq 0$, $c_5 > 0$ and $c_6 \geq 0$. The latter inequality holds due to the positivity of F and due to the monotone decrease of the sequence $\{F(u_k)\}$ with the corresponding choice of the damping parameters $\{\omega_k\}$ in (5.58). For this reason, the sequence $\{u_k\}$ is uniformly bounded in the BV norm and the Banach-Alaoglu theorem delivers the precompactness in the weak* topology on $BV(\Omega)$, which implies the existence of a subsequence $\{u_{k_l}\}$ with

$$u_{k_l} \;\rightharpoonup^*\; u \qquad \text{in}\;\; BV(\Omega)\;.$$

The definition of the weak* topology on $BV(\Omega)$ in [4, Definition 3.11] implies that the subsequence $\{u_{k_l}\}$ also converges strongly in $L^1(\Omega)$,

$$u_{k_l} \;\to\; u \qquad \text{in} \;\; L^1(\Omega) \; .$$

With the same argumentation, we can choose further subsequences, again denoted by k_l, such that

$$\begin{aligned} u_{k_l+1} \;&\overset{*}{\rightharpoonup}\; \tilde{u} \qquad \text{in} \;\; BV(\Omega) \; , \\ u_{k_l+1} \;&\to\; \tilde{u} \qquad \text{in} \;\; L^1(\Omega) \; . \end{aligned}$$

We show that the limits of the subsequences $\{u_{k_l}\}$ respectively $\{u_{k_l+1}\}$ are the same, i.e. $u \;=\; \tilde{u}$. For this purpose, we apply inequality (5.60) recursively and obtain the following estimate,

$$F(u_{k+1}) \;+\; \epsilon \sum_{j\,=\,0}^{k} \int_\Omega \frac{K^*\mathbf{1}_\Sigma \,(u_{j+1} \;-\; u_j)^2}{\omega_j \; u_j} \; d\lambda \;\leq\; F(u_0) \;<\; \infty \; , \qquad \forall k \;\geq\; 0 \; .$$

Thus, the series of the damping terms is summable and the Cauchy criterion for convergence delivers

$$\lim_{k\to\infty} \int_\Omega \frac{K^*\mathbf{1}_\Sigma \,(u_{k+1} \;-\; u_k)^2}{\omega_k \; u_k} \; d\lambda \;=\; 0 \; . \tag{5.65}$$

Additionally, the Cauchy-Schwarz inequality yields the following estimate,

$$\|u_{k+1} \;-\; u_k\|_1^2 \;\leq\; \int_\Omega \frac{\omega_k \; u_k}{K^*\mathbf{1}_\Sigma} \; d\lambda \; \int_\Omega \frac{K^*\mathbf{1}_\Sigma \,(u_{k+1} \;-\; u_k)^2}{\omega_k \; u_k} \; d\lambda \; . \tag{5.66}$$

The first term on the right-hand side of (5.66) is uniformly bounded for all $k \geq 0$, since $\omega_k \;\in\; (0,1]$, the function $K^*\mathbf{1}_\Sigma$ is bounded away from zero (5.59) and the sequence $\{u_k\}$ is uniformly bounded in the BV norm. From (5.65), the second term converges to zero and we obtain from (5.66) that

$$u_{k+1} \;-\; u_k \;\to\; 0 \qquad \text{in} \;\; L^1(\Omega) \; . \tag{5.67}$$

Hence, the uniqueness of the limit implies $u \;=\; \tilde{u}$.

Third step: Convergence of the dual iterates

In addition we need a convergent subsequence of the subgradients $\{p_k\}$ corresponding to the sequence $\{u_k\}$, i.e. $p_k \in \partial \, |u_k|_{BV(\Omega)}$. For this sake, we use the general property that the subdifferential of a convex one-homogeneous functional $J : X \to \mathbb{R} \cup \{+\infty\}$, X Banach space, can be characterized by

$$\partial J(u) \;=\; \{\, p \in X^* \;:\; \langle p, u\rangle \;=\; J(u) \; , \quad \langle p, v\rangle \;\leq\; J(v) \quad \forall v \in X \,\} \; .$$

In the case of TV, we see that for each subgradient p_k the dual norm is bounded by

$$\|p_k\| \;=\; \sup_{\|v\|_{BV(\Omega)}=1} \langle p_k, v\rangle \;\leq\; \sup_{\|v\|_{BV(\Omega)}=1} |v|_{BV(\Omega)} \;\leq\; \sup_{\|v\|_{BV(\Omega)}=1} \|v\|_{BV(\Omega)} \;=\; 1\ .$$

Hence, the sequence $\{p_k\}$ is uniformly bounded in the BV* norm and the Banach-Alaoglu theorem delivers the precompactness in the weak* topology on $(BV(\Omega))^*$, which implies the existence of a subsequence, again denoted by k_l, such that

$$p_{k_l+1} \;\rightharpoonup^*\; p \qquad \text{in}\ \ (BV(\Omega))^*\ .$$

Fourth step: Verification of $p \in \partial\,|u|_{BV(\Omega)}$
We have now the weak* convergence of u_{k_l} respectively u_{k_l+1} in $BV(\Omega)$ and the weak* convergence of p_{k_l+1} in $(BV(\Omega))^*$. Next, we will show that the limit p of the dual iterates is a subgradient of $|\cdot|_{BV(\Omega)}$ at the limit u of the primal iterates, i.e. $p \in \partial\,|u|_{BV(\Omega)}$. Hence we have to prove, see the definition of the subdifferential in (3.4), that

$$|u|_{BV(\Omega)} \;+\; \langle p, v-u\rangle \;\leq\; |v|_{BV(\Omega)}\ , \qquad \forall v \in BV(\Omega)\ .$$

Let $p_{k_l+1} \in \partial\,|u_{k_l+1}|_{BV(\Omega)}$, then and the definition of the subdifferential of $|\cdot|_{BV(\Omega)}$ (3.4) yields

$$|u_{k_l+1}|_{BV(\Omega)} \;+\; \langle p_{k_l+1}, v-u_{k_l+1}\rangle \;\leq\; |v|_{BV(\Omega)}\ , \qquad \forall v \in BV(\Omega)\ . \tag{5.68}$$

Since $|\cdot|_{BV(\Omega)}$ is lower semicontinuous we can estimate the BV seminorm of the primal iterates from below,

$$|u|_{BV(\Omega)} \;\leq\; \liminf_{l\to\infty} |u_{k_l+1}|_{BV(\Omega)} \;\leq\; |u_{k_l+1}|_{BV(\Omega)}\ ,$$

and (5.68) delivers

$$|u|_{BV(\Omega)} \;+\; \langle p_{k_l+1}, v-u_{k_l+1}\rangle \;\leq\; |v|_{BV(\Omega)}\ , \qquad \forall v \in BV(\Omega)\ . \tag{5.69}$$

In addition, in the previous step we verified the weak* convergence of $\{p_{k_l+1}\}$ in $(BV(\Omega))^*$, i.e.

$$\langle p_{k_l+1}, v\rangle \;\to\; \langle p, v\rangle\ , \qquad \forall v \in BV(\Omega)\ .$$

Hence, in order to prove $p \in \partial\,|u|_{BV(\Omega)}$, it remains to with respect to (5.69) that

$$\langle p_{k_l+1}, -u_{k_l+1}\rangle \;\to\; \langle p, -u\rangle\ . \tag{5.70}$$

For this purpose we consider the complete iteration scheme of the damped FB-EM-TV algorithm, see (5.27) and (5.25), concerning the convergent subsequences,

$$u_{k_l+1} \;-\; \omega_{k_l}\left(u_{k_l}\,\frac{K^*}{K^*\mathbf{1}_\Sigma}\left(\frac{f}{Ku_{k_l}}\right)\right) \;-\; (1-\omega_{k_l})\,u_{k_l} \;+\; \omega_{k_l}\,\alpha\,\frac{u_{k_l}}{K^*\mathbf{1}_\Sigma}\,p_{k_l+1} \;=\; 0\ , \tag{5.71}$$

which is equivalent to

$$-\alpha\, p_{k_l+1} \;=\; \frac{K^*\mathbf{1}_\Sigma\,(u_{k_l+1} \,-\, u_{k_l})}{\omega_{k_l}\, u_{k_l}} \;+\; K^*\mathbf{1}_\Sigma \;-\; K^*\left(\frac{f}{Ku_{k_l}}\right) .$$

Multiplying this formulation of the iteration scheme with u_{k_l+1} and integrating over domain Ω yields

$$\begin{aligned}
&-\alpha\,\langle p_{k_l+1}, u_{k_l+1}\rangle \\
&= \int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k_l+1} \,-\, u_{k_l})\,u_{k_l+1}}{\omega_{k_l}\, u_{k_l}}\, d\lambda \;+\; \langle \mathbf{1}_\Sigma, Ku_{k_l+1}\rangle \;-\; \left\langle \frac{f}{Ku_{k_l}}, Ku_{k_l+1}\right\rangle \\
&= \underbrace{\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k_l+1} \,-\, u_{k_l})^2}{\omega_{k_l}\, u_{k_l}}\, d\lambda}_{\overset{(5.65)}{\longrightarrow}\, 0} \;+\; \underbrace{\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k_l+1} \,-\, u_{k_l})\,u_{k_l}}{\omega_{k_l}\, u_{k_l}}\, d\lambda}_{\longrightarrow\, 0} \qquad (5.72)\\
&\qquad\qquad +\; \langle \mathbf{1}_\Sigma, Ku_{k_l+1}\rangle \;-\; \left\langle \frac{f}{Ku_{k_l}}, Ku_{k_l+1}\right\rangle .
\end{aligned}$$

The second term on the right-hand side of (5.72) vanishes in the limit since $\frac{K^*\mathbf{1}_\Sigma}{\omega_{k_l}}$ is uniformly bounded in the supremum norm, due to the boundedness away from zero of ω_k (5.59) and the assumption (A4), and the convergence (5.67). Furthermore, using the boundedness of the operator K for the convergence of $\frac{f}{Ku_{k_l}}$, we obtain

$$\langle \mathbf{1}_\Sigma, Ku_{k_l+1}\rangle \;-\; \left\langle \frac{f}{Ku_{k_l}}, Ku_{k_l+1}\right\rangle \;\to\; \langle \mathbf{1}_\Sigma, Ku\rangle \;-\; \left\langle \frac{f}{Ku}, Ku\right\rangle$$

and can deduce from (5.72) that

$$-\alpha\,\langle p_{k_l+1}, u_{k_l+1}\rangle \;\to\; \int_\Omega \left(K^*\mathbf{1}_\Sigma \;-\; K^*\left(\frac{f}{Ku}\right)\right) u\; d\lambda \;\overset{(5.74)}{=}\; -\alpha\,\langle p, u\rangle .$$

Hence, we can conclude (5.70) and that $p \in \partial\,|u|_{BV(\Omega)}$.

Fifth step: Convergence to a minimizer of the functional

We consider the complete iteration scheme of the damped FB-EM-TV algorithm (5.71) with respect to the convergent subsequences and show the convergence in the weak* topology to the optimality condition (5.20) of the variational problem (5.35). An equivalent formulation of equation (5.71) reads as follows

$$\frac{u_{k_l+1} \,-\, u_{k_l}}{\omega_{k_l}\, u_{k_l}} \;+\; 1 \;-\; \frac{K^*}{K^*\mathbf{1}_\Sigma}\left(\frac{f}{Ku_{k_l}}\right) \;+\; \frac{\alpha}{K^*\mathbf{1}_\Sigma}\, p_{k_l+1} \;=\; 0 . \qquad (5.73)$$

The convergence can be verified in the following way. Due to the assumptions of the boundedness away from zero in (5.59), we can use result (5.65) in order to deduce the following convergence

$$c_1\, c_2\, c_3 \int_\Omega \frac{(u_{k+1} \,-\, u_k)^2}{\omega_k^2\, u_k^2}\, d\lambda \;\le\; \int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u_{k+1} \,-\, u_k)^2}{\omega_k^2\, u_k^2}\, \omega_k\, u_k\, d\lambda \;\overset{(5.65)}{\longrightarrow}\; 0 .$$

Since the integrand on the left-hand side is positive, we obtain with the uniqueness of the limit, that

$$\lim_{l \to \infty} \frac{u_{k_l+1} - u_{k_l}}{\omega_{k_l} u_{k_l}} = 0 .$$

Therefore, if we pass over to the weak* limit of the subsequences in (5.73), using the boundedness of the operator K for the convergence of $\frac{f}{Ku_{k_l}}$, the function u and the subgradient $p \in \partial |u|_{BV(\Omega)}$ fulfill the optimality condition (5.20) of the variational problem (5.35),

$$1 - \frac{K^*}{K^*\mathbf{1}_\Sigma} \left(\frac{f}{Ku} \right) + \alpha \frac{1}{K^*\mathbf{1}_\Sigma} p = 0 . \tag{5.74}$$

This means that the subsequence $\{u_{k_l}\}$ converges in the weak* topology on $BV(\Omega)$ and in the strong topology on $L^1(\Omega)$ to a minimizer of the functional F. Note that it suffices to show (5.74) instead of (5.24) since with the assumption $0 < c_3 \leq u_k$ for all $k \geq 0$ in (5.59) is u automatically positive with $0 < c_3 \leq u$.

Since we assumed the operator K to be injective, we obtain the strong convexity of the functional F from Theorem 5.4.8 and for this reason a unique minimizer u. This means that every convergent subsequence has the same limit. Thus, also the sequence $\{u_k\}$ converges to the unique minimizer in the weak* topology on $BV(\Omega)$ and in the strong topology on $L^1(\Omega)$. □

Remarks and Extensions. We mention that the equation (5.64) motivated at the same time the necessity of a damping in the FB-EM-TV splitting strategy. In the undamped case, i.e. $\omega_k = 1$, the condition (5.64) is in general not fulfilled and hence we need a parameter $\omega_k \in (0, 1)$ increasing the term on the right-hand side of (5.64) in order to guarantee a descent of the objective functional F.

Analogous to above, the strong convergence of the sequence $\{u_k\}$ to a minimizer of the functional F in Theorem 5.4.12 can also be proved in the L^p norm with $1 \leq p < d/(d-1)$, since any uniformly bounded sequence $\{u_k\}$ in $BV(\Omega)$ is actually relatively compact in $L^p(\Omega)$ for $1 \leq p < d/(d-1)$, see [1, Theorem 2.5]. Therefore, since the subsequence $\{u_{k_l}\}$ is furthermore uniformly bounded in the BV norm, there exists a subsequence $\{u_{k_{l_m}}\}$ with

$$u_{k_{l_m}} \to \tilde{u} \qquad \text{in } L^p(\Omega) \text{ with } 1 \leq p < d/(d-1) .$$

With the uniqueness of the limit and the definition of the weak* topology on $BV(\Omega)$ we obtain

$$\begin{aligned} u_{k_{l_m}} &\rightharpoonup^* u \qquad \text{in } BV(\Omega) , \\ u_{k_{l_m}} &\to u \qquad \text{in } L^1(\Omega) . \end{aligned}$$

Due to the uniqueness of the limit, i.e. $\tilde{u} = u$, we can pass over in the proof from $\{u_{k_l}\}$ to $\{u_{k_{l_m}}\}$.

The assumptions on boundedness away from zero in (5.59) are reasonable. Let us consider the case of the function $K^*\mathbf{1}_\Sigma$. The assumption is practical, because if there is a point $x \in \Omega$ with $(K^*\mathbf{1}_\Sigma)(x) = 0$, then it is a priori impossible to reconstruct the information in this point. Moreover, the assertion on the damping parameters makes sense because a strong damping is certainly undesirable. The boundedness of the iterates u_k is fulfilled due to the strict positivity of each halfstep of the FB-EM-TV splitting method, see Lemma 5.4.11.

Inspired by the relaxed EM reconstruction strategy proposed in [133, Chapter 5.3.2], another possibility of influencing convergence arises in FB-EM-TV by adding a relaxation parameter $\nu > 0$ to the EM fixed point iteration,

$$u_{k+\frac{1}{2}} \;=\; u_k \left(\frac{K^*}{K^*\mathbf{1}_\Sigma} \left(\frac{f}{K u_k} \right) \right)^\nu \qquad \text{(relaxed EM step)} \; .$$

One can obtain a reasonable TV denoising step in the FB-EM-TV splitting strategy via

$$u_{k+1} \;=\; \left(u_{k+\frac{1}{2}}^{\frac{1}{\nu}} \;-\; \alpha \, u_k^{\frac{1}{\nu}} \, p_{k+1} \right)^\nu \qquad \text{(relaxed TV step)} \; ,$$

with $p_{k+1} \in \partial \, |u_{k+1}|_{BV(\Omega)}$, corresponding to the relaxed EM step $u_{k+\frac{1}{2}}$ above. The relaxed terms in the TV denoising step are necessary to fit the basic variational problem (5.18), respectively its corresponding optimality condition. Due to the computational challenge of the relaxed TV denoising step, which would require again novel methods, a comparison of this strategy with our damping strategy would go beyond the scope of this thesis.

In practice, determining the damping parameters ω_k via the general condition in (5.58) is not straight-forward and one can be interested in an explicit bound for all damping parameters ω_k. That is why we will provide an explicit bound on ω_k in the case of Poisson denoising, i.e. $K = Id$. In this case, the EM reconstruction step in (5.25) reduces to $u_{k+\frac{1}{2}} = f$ and we compute only,

$$u_{k+1} \;=\; \underset{u \, \in \, BV(\Omega)}{\arg\min} \left\{ \frac{1}{2} \int_\Omega \frac{(u \;-\; (\omega_k \; f \;+\; (1-\omega_k) \, u_k))^2}{u_k} \;+\; \omega_k \, \alpha \, |u|_{BV(\Omega)} \right\} \; . \tag{5.75}$$

In Section 6.1 we will see that a reasonable choice of such a bound can improve the convergence behavior significantly. For the denoising scheme a maximum principle will be a key step to provide an explicit bound on ω_k.

Lemma 5.4.13 (Maximum Principle for $K = Id$)**.** *Let $\{\omega_k\}$ be a sequence of damping parameters with $\omega_k \in (0, 1]$ for all $k \geq 0$ and the initialization function u_0 fulfill*

$$0 \;<\; \min f \;\leq\; \min u_0 \;\leq\; \max u_0 \;\leq\; \max f \;. \tag{5.76}$$

Moreover, let $\{u_k\}$ be a sequence of iterates generated by the damped Poisson denoising scheme (5.75). Then, the following maximum principle holds

$$0 \;<\; \min f \;\leq\; \min u_k \;\leq\; \max u_k \;\leq\; \max f \;, \qquad \forall k \;\geq\; 0 \;. \tag{5.77}$$

Proof. We prove the assertion by induction. For $k = 0$, the condition (5.77) is fulfilled due to (5.76). For a general $k \geq 0$, Lemma 5.4.10 offers a maximum principle for the weighted ROF denosing model (5.75),

$$\begin{aligned} 0 \;<\; \min\{\, \omega_k \; f \;+\; (1-\omega_k)\, u_k \,\} \;&\leq\; \min u_{k+1} \\ &\leq\; \max u_{k+1} \;\leq\; \max\{\, \omega_k \; f \;+\; (1-\omega_k)\, u_k \,\} \;. \end{aligned} \tag{5.78}$$

Due to the fact that $\omega_k \in (0, 1]$ for all $k \geq 0$ and the inequalities

$$\min\{\, \omega_k \; f \;+\; (1-\omega_k)\, u_k \,\} \;\geq\; \omega_k \; \min f \;+\; (1-\omega_k) \; \min u_k$$

and

$$\max\{\, \omega_k \; f \;+\; (1-\omega_k)\, u_k \,\} \;\leq\; \omega_k \; \max f \;+\; (1-\omega_k) \; \max u_k \;,$$

we obtain from (5.78) and the induction hypothesis the desired maximum principle (5.77). □

Remarks and Extensions. The assumption (5.76) is fulfilled in general, since the initialization u_0 will be usually chosen as a positive and constant function.

Corollary 5.4.14 (Explicit Bound on ω_k for $K = Id$)**.** *Let $\{u_k\}$ be a sequence of iterates generated by the damped Poisson denoising scheme (5.75). To guarantee the convergence in the case of $K = Id$, the condition (5.58) in Theorem 5.4.12 on the damping parameters simplifies to*

$$\omega_k \;\leq\; \frac{2\,(\min f)^2}{(\max f)^2}\,(1-\epsilon) \;, \qquad \epsilon \;\in\; (0,1) \;. \tag{5.79}$$

Proof. In the special case of $K = Id$, the maximum principle from Lemma 5.4.13 is the key idea for simplifying the condition (5.58) on the damping parameters. For this sake, we consider the inequality (5.64), according to which a monotone descent of the objective functional is guaranteed if

$$\frac{1}{2}\int_\Omega \frac{f\,u_k}{v^2}\,\frac{(u_{k+1} \;-\; u_k)^2}{u_k} \;\leq\; \frac{1-\epsilon}{\omega_k}\int_\Omega \frac{(u_{k+1}-u_k)^2}{u_k} \;, \qquad \forall\, v \;\in\; [u_k, u_{k+1}] \;,$$

holds. Our goal is now to find an estimate for the coefficients $\frac{f u_k}{2v^2}$. Due to the fact that $v \in [u_k, u_{k+1}]$ and that $\{u_k\}$ are iterates generated by the damped Poisson denoising scheme (5.75), we can use the maximum principle from Lemma 5.4.13 and obtain an estimate for the coefficients,

$$\frac{f\, u_k}{2\, v^2} \;\leq\; \frac{\max f \;\max u_k}{2\,(\min\{u_k,\; u_{k+1}\})^2} \;\leq\; \frac{(\max f)^2}{2\,(\min f)^2}\;, \qquad \forall\, k \;\geq\; 0\;,$$

which should be less or equal $\frac{1-\epsilon}{\omega_k}$. Thus, choosing ω_k according to the estimate (5.79) guarantees a monotone descent of the objective functional. □

5.5 Numerical Realization of weighted ROF

In this section we discuss the numerical realization of the TV correction half step which occurs in the (damped) FB-EM-TV and the (damped) Bregman-EM-TV algorithm, proposed in Chapter 5.3. In the latter, we suggested to realize the TV denoising steps equivalently by the solution of the weighted ROF problems, see (5.26), (5.28), (5.93) and (5.94). The most general form of all the schemes above is

$$u \;=\; \operatorname*{arg\,min}_{u\,\in\,BV(\Omega)} \left\{ \frac{1}{2} \int_\Omega \frac{(u \,-\, q)^2}{h} \;+\; \beta\, |u|_{BV(\Omega)} \right\}\;, \qquad \beta > 0\;, \tag{5.80}$$

with an appropriate setting of the noise function q, the weight function h and the regularization parameter β, namely

K = Id (Poisson Denoising)

$$q \;:=\; f\;, \qquad h \;:=\; u_k\;, \qquad \beta \;:=\; \alpha$$

FB-EM-TV Algorithm (5.26)

$$q \;:=\; u_{k+\frac{1}{2}}\;, \qquad h \;:=\; \frac{u_k}{K^*\mathbf{1}_\Sigma}\;, \qquad \beta \;:=\; \alpha$$

Damped FB-EM-TV Algorithm (5.28)

$$q \;:=\; \omega_k\, u_{k+\frac{1}{2}} \;+\; (1-\omega_k)\, u_k\;, \qquad h \;:=\; \frac{u_k}{K^*\mathbf{1}_\Sigma}\;, \qquad \beta \;:=\; \omega_k\, \alpha$$

Bregman-EM-TV Algorithm (5.93)

$$q \;:=\; u_{k+\frac{1}{2}}^{l+1} \;+\; u_k^{l+1}\, v^l\;, \qquad h \;:=\; \frac{u_k^{l+1}}{K^*\mathbf{1}_\Sigma}\;, \qquad \beta \;:=\; \alpha$$

Damped Bregman-EM-TV Algorithm (5.94)

$$q \;:=\; \omega_k^{l+1}\, u_{k+\frac{1}{2}}^{l+1} + (1-\omega_k^{l+1})\, u_k^{l+1} + \omega_k^{l+1}\, u_k^{l+1}\, v^l\;, \qquad h \;:=\; \frac{u_k^{l+1}}{K^*\mathbf{1}_\Sigma}\;, \qquad \beta \;:=\; \omega_k^{l+1}\, \alpha$$

5.5.1 Dual Implementation

For the original ROF model, i.e. $q := f$, $h := \mathbf{1}_\Omega$ and $\beta := \alpha$, a variety of minimization approaches are known in literature, e.g. we refer to [42], [43], [9]. Most of these computational schemes can be adapted to the weighted modification (5.80). Here, we use for the minimization the exact dual approach (5.6), which does not need any smoothing of the total variation. Our approach is analogous to the one of Chambolle in [42], which characterizes the subgradients of TV as divergences of vector fields with supremum norm less or equal one and allows an accurate, robust and efficient algorithm.

In the following, we establish an iteration to compute the solution of the variational problem (5.80). This minimization can be written as a saddle point problem in the primal variable u and the dual variable g,

$$\inf_{u \in BV(\Omega)} \sup_{\substack{g \in C_0^\infty(\Omega, R^d) \\ ||g||_\infty \leq 1}} \left\{ L(u,g) \;:=\; \frac{1}{2} \int_\Omega \frac{(u - q)^2}{h} \;+\; \beta \int_\Omega u \, \nabla \cdot g \right\} . \tag{5.81}$$

Formally, the infimum at u and the supremum at p can be changed. In the case of the standard ROF model, i.e if the weight h in (5.80) is missing, this property is proved in [129], which can be carried over with minimal modifications. Moreover, a more precise analysis of this property for general saddle point problems is available in [62, Proposition 2.3, p. 175]. After exchange of inf and sup, the primal optimality condition for the saddle point problem (5.81) is given by

$$\frac{\partial}{\partial u} L(u,g) \;=\; 0 \qquad \Leftrightarrow \qquad u \;=\; q \;-\; \beta \, h \, \nabla \cdot g \; . \tag{5.82}$$

If the optimal dual variable $\tilde{g}$ is known, the condition (5.82) can be used to obtain a solution of (5.81),

$$u \;=\; q \;-\; \beta \, h \, \nabla \cdot \tilde{g} \; . \tag{5.83}$$

For the computation of $\tilde{g}$, we plug (5.82) into (5.81) and obtain a purely dual problem which depends on g only. With terms that are constant with respect to the optimization variable and hence do not change the supremum, and under the substitution of the maximization by minimization of the negative functional, we obtain

$$\begin{aligned} \tilde{g} \;=\; & \operatorname*{arg\,min}_{g \in C_0^\infty(\Omega, R^d)} \int_\Omega \frac{(\beta \, h \, \nabla \cdot g \;-\; q)^2}{h} \\ & \text{subject to} \;\; |g(x)|_{\ell^2}^2 \;-\; 1 \;\leq\; 0 \; , \qquad \forall x \in \Omega \; , \end{aligned} \tag{5.84}$$

where $|\cdot|_{\ell^2}$ is the Euclidean vector norm. For the choice of the vector norm, compare the remark at the end of this chapter. The constraint for the dual variable g in (5.84) is a consequence of the exact (dual) definition of total variation (5.6).

The dual problem (5.84) is a (weighted) quadratic optimization problem with a non-linear inequality constraint. Hence, for the computation of the optimal dual variable $\tilde{g}$, we use the Karush-Kuhn-Tucker (KKT) conditions (cf. e.g. [100, Theorem 2.1.4, p. 305]), which yields the existence of a Lagrange multiplier $\lambda(x) \geq 0$ for almost every $x \in \Omega$ such that

$$-\nabla(\beta\, h\, \nabla \cdot g \;-\; q)(x) \;+\; \lambda(x)\, g(x) \;=\; 0\;, \qquad \forall x \in \Omega\;, \tag{5.85}$$

and

$$\lambda(x)\,(\,|g(x)|_{\ell^2}^2 \;-\; 1\,) \;=\; 0\;, \qquad \forall x \in \Omega\;. \tag{5.86}$$

The multiplier λ can be specified explicitly from the complementarity condition (5.86), which yields $\lambda(x) > 0$ and $|g(x)|_{\ell^2} = 1$, or $\lambda(x) = 0$. Thus, in any case we obtain from (5.85),

$$\lambda(x) \;=\; |\lambda(x)\, g(x)|_{\ell^2} \;=\; |\nabla(\beta\, h\, \nabla \cdot g \;-\; q)(x)|_{\ell^2}\;, \qquad \forall x \in \Omega\;.$$

We can write (5.85) as a fixed point equation for g and iterate

$$g^{n+1}(x) \;=\; \frac{g^n(x) \;+\; \tau\,(\nabla(\beta\, h\, \nabla \cdot g^n \;-\; q)(x))}{1 \;+\; \tau\,|\nabla(\beta\, h\, \nabla \cdot g^n \;-\; q)(x)|_{\ell^2}}\;, \qquad \forall x \in \Omega\;. \tag{5.87}$$

In a standard discrete setting on pixels with unit size and first derivatives computed by one-sided differences the convergence result in [42, Theorem 3.1] can be transferred to the weighted ROF problem (5.80), such that we can guarantee the convergence to a solution if the damping parameter τ satisfies

$$0 \;<\; \tau \;\leq\; \frac{1}{4\, d\, \beta\, ||h||_\infty}\;.$$

The weight h in (5.80) can also be interpreted as an adaptive regularization, because the regularization parameter β is weighted in (5.87) by the function h.

Remarks and Extensions. Finally, we point out that the total variation definition (5.6) is not unique for $d > 1$. Depending on the definition of the supremum norm $||g||_\infty = \operatorname{ess\,sup}_{x\in\Omega} |g(x)|_{\ell^p}$ for $g \in C_0^\infty(\Omega, R^d)$ and $1 \leq p < \infty$, we obtain a family of total variation seminorms. The most common formulations are the isotropic total variation ($p = 2$) that we used here, see the minimization constraint in (5.84), and the anisotropic total variation ($p = \infty$). Different definitions of TV have effects on the structure of minimizers of (5.80). In the case of the isotropic TV, corners in the edge set will not be allowed, whereas orthogonal corners are favored by the anisotropic variant. For a detailed analysis, we e.g. refer to [124], [171], [67], [21].

5.6 Contrast Enhancement via Bregman Iterations

In this section we apply the idea of primal Bregman iteration and the new idea of dual Bregman iteration to reconstruction problems with non-standard noise models. Analog to the previous sections we on 3D imaging we concentrate on the case of Poisson noise. This means, we have the Kullback-Leibler distance as the underlying data fidelity and consider Bregman distance regularization regarding total variation.

5.6.1 Primal Bregman-EM-TV

The presented FB-EM-TV algorithm (5.25) solves the problem (5.18) and delivers cartoon-like reconstructions with sharp edges. However, the realization of TV correction steps via the weighted ROF model (5.26) has the drawback that reconstructed images suffer from contrast reduction [124], [134]. Thus, we propose to extend (5.18) and therewith FB-EM-TV by iterative regularization to a simultaneous contrast correction. More precisely, we perform a constrast enhancement by inverse scale space methods and by using the Bregman iteration. These techniques have been derived in [134], with a detailed analysis for Gaussian-type problems (5.4), and have been generalized to time-continuity [34] and L^p-norm data fitting terms [33]. Following these methods, an iterative refinement is realized by a sequence of modified EM-TV problems based on (5.18).

The inverse scale space methods concerning TV, derived in [134], follow the concept of iterative regularization by Bregman distance [27]. In case of the Poisson model, the method initially starts with a simple FB-EM-TV algorithm, i.e. it consists in computing a minimizer u^1 of (5.18), respectively (5.19). Then, updates are determined successively by considering variational problems with a shifted TV term,

$$u^{l+1} \;=\; \underset{\substack{u \in BV(\Omega) \\ u \geq 0 \;\; a.e.}}{\arg\min} \left\{ \; D_{KL}(f, Ku) \;+\; \alpha \, \left(\, |u|_{BV(\Omega)} \;-\; \langle p^l, u \rangle \, \right) \; \right\} \;, \tag{5.88}$$

where $p^l \in \partial \, |u^l|_{BV(\Omega)}$ and $\alpha > 0$. The mentioned Bregman distance with respect to $|\cdot|_{BV(\Omega)}$ is defined via

$$D^{\tilde{p}}_{|\cdot|_{BV(\Omega)}}(u, \tilde{u}) \;:=\; |u|_{BV(\Omega)} - |\tilde{u}|_{BV(\Omega)} - \langle \tilde{p}, u - \tilde{u} \rangle \;, \qquad \tilde{p} \in \partial \, |\tilde{u}|_{BV(\Omega)} \subseteq (BV(\Omega))^* \;,$$

where $\langle \cdot, \cdot \rangle$ denotes the standard duality product. The introduction of this definition allows to characterize the sequence of modified variational problems (5.88) by adding of constant terms as

$$u^{l+1} \;=\; \underset{\substack{u \in BV(\Omega) \\ u \geq 0 \;\; a.e.}}{\arg\min} \left\{ \; D_{KL}(f, Ku) \;+\; \alpha \, D^{p^l}_{|\cdot|_{BV(\Omega)}}(u, u^l) \right\}, \qquad p^l \in \partial \, |u^l|_{BV(\Omega)} \;. \tag{5.89}$$

The first iterate u^1 can also be realized by the variational problem (5.88) or (5.89), if u^0 is constant and $p^0 := 0 \in \partial\,|u^0|_{BV(\Omega)}$.

The Bregman distance $D^{\tilde{p}}_{|\cdot|_{BV(\Omega)}}$ does not represent a distance in the common (metric) sense, since D is not symmetric in general and the triangle inequality does not hold. However, the formulation in (5.89) has the advantage over (5.88), that $D^{\tilde{p}}_{|\cdot|_{BV(\Omega)}}$ is a distance measure in the following sense

$$D^{\tilde{p}}_{|\cdot|_{BV(\Omega)}}(u,\tilde{u}) \;\geq\; 0 \qquad \text{and} \qquad D^{\tilde{p}}_{|\cdot|_{BV(\Omega)}}(u,\tilde{u}) \;=\; 0 \quad \text{for} \;\; u \;=\; \tilde{u}\;.$$

Besides, the Bregman distance is convex in the first argument because $|\cdot|_{BV(\Omega)}$ is convex. In general, i.e. for any convex functional J, see e.g. [33], the Bregman distance can be interpreted as the difference between J in u and the Taylor linearization of J around $\tilde{u}$ if, in addition, J is continuously differentiable.

From the point of view of the statistical problem formulation in Section 5.2.2, the Bregman regularized variational problem (5.89) uses in (2.6) an adapted a-priori probability density $p(u)$. Instead of a zero-centered a-priori probability with $J(u) = |u|_{BV(\Omega)}$ as in case of the FB-EM-TV algorithm, here we consider in every Bregman refinement step a new a-priori probability which is related to a shifted total variation, i.e. we use the Gibbs function (2.6) with

$$J(u) \;=\; D^{p^l}_{|\cdot|_{BV(\Omega)}}(u,u^l)\;.$$

This means that images with smaller total variation and a close distance to the maximum likelihood estimator u^l of the previous FB-EM-TV problem are preferred in the minimization (5.89).

Before deriving a two-step iteration corresponding to (5.25), we will motivate the contrast enhancement by iterative regularization in (5.89). The regularization with TV in (5.19) prefers functions with only few oscillations, ideally piecewise constant functions. As a consequence, the reconstruction results suffer from a loss of contrast. The iterative Bregman regularization has the advantage that, with u^l as an approximation to the possible solution, additional information is available. The variational problem (5.89) can be interpreted as follows: search for a solution that matches the Poisson distributed data after applying K and simultaneous minimizes the residual of the Taylor approximation of $|\cdot|_{BV(\Omega)}$ around u^l. This form of regularization hardly changes the position of gradients with respect to the last iterate u^l, but that an increase of intensities is permitted. This leads to a noticeable contrast enhancement.

For designing a two-step iteration analogous to FB-EM-TV algorithm, we consider the first order optimality condition for the variational problem (5.89). Due to the

convexity of the Bregman distance in the first argument, we can determine the subdifferential of (5.89). Analogous to the derivation of the FB-EM-TV iteration, due to the continuity of the Kullback-Leibler functional, we can split the subdifferential of a sum of functions to a sum of subdifferentials and the subdifferential of the KL functional can be expressed formally by the Fréchet derivatives like to (5.20). Since

$$\partial\,(\,-\,\langle p^l, u\rangle\,)\;=\;\{\,-p^l\,\}$$

holds, the first order optimality condition of (5.89) for a positive solution u^{l+1} is given by

$$0\;\in\;K^*\mathbf{1}_\Sigma - K^*\left(\frac{f}{Ku^{l+1}}\right) + \alpha\,(\,\partial\,|u^{l+1}|_{BV(\Omega)} - p^l\,)\,,\qquad p^l\;\in\;\partial\,|u^l|_{BV(\Omega)}\,. \tag{5.90}$$

For u^0 constant and $p^0 := 0 \in \partial\,|u^0|_{BV(\Omega)}$ this condition delivers a well defined update of the iterates p^l,

$$p^{l+1}\;:=\;p^l\;-\;\frac{1}{\alpha}\left(\,K^*\mathbf{1}_\Sigma\;-\;K^*\left(\frac{f}{Ku^{l+1}}\right)\right)\;\in\;\partial\,|u^{l+1}|_{BV(\Omega)}\,.$$

Analogous to the FB-EM-TV algorithm, we can apply the idea of the nested iteration (5.25) in every refinement step, $l = 1, 2, \cdots$. Thus, for the solution of (5.89), the condition (5.90) yields a strategy consisting of an EM-step $u^{l+1}_{k+\frac{1}{2}}$ followed by solving the adapted weighted ROF problem

$$u^{l+1}_{k+1}\;=\;\underset{u\,\in\,BV(\Omega)}{\arg\min}\left\{\;\frac{1}{2}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u\;-\;u^{l+1}_{k+\frac{1}{2}})^2}{u^{l+1}_k}\;+\;\alpha\,(\,|u|_{BV(\Omega)}\;-\;\langle p^l, u\rangle\,)\;\right\}\,. \tag{5.91}$$

Following [134], [34], [33], we provide an opportunity to transfer the shift term $\langle p^l, u\rangle$ to the data fidelity term. This approach facilitates the implementation of contrast enhancement with Bregman distance via a slightly modified FB-EM-TV algorithm. With the scaling $K^*\mathbf{1}_\Sigma\,v^l := \alpha\,p^l$ and (5.90) we obtain the following update formula

$$v^{l+1}\;=\;v^l\;-\;\left(\,1\;-\;\frac{K^*}{K^*\mathbf{1}_\Sigma}\left(\frac{f}{Ku^{l+1}}\right)\right)\,,\qquad v^0\;=\;0\,. \tag{5.92}$$

Using this scaled update we can rewrite the second step (5.91) to

$$u^{l+1}_{k+1}\;=\;\underset{u\,\in\,BV(\Omega)}{\arg\min}\left\{\;\frac{1}{2}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(\,(u\;-\;u^{l+1}_{k+\frac{1}{2}})^2\;-\;2uu^{l+1}_k v^l\,)}{u^{l+1}_k}\;+\;\alpha\,|u|_{BV(\Omega)}\;\right\}\,.$$

Note that

$$(u\;-\;u^{l+1}_{k+\frac{1}{2}})^2\;-\;2uu^{l+1}_k v^l\;=\;(\,u\;-\;(u^{l+1}_{k+\frac{1}{2}}\;+\;u^{l+1}_k v^l)\,)^2\;-\;2u^{l+1}_{k+\frac{1}{2}}u^{l+1}_k v^l\;+\;(u^{l+1}_k)^2(v^l)^2$$

holds, where the last two terms are independent of u. Hence (5.91) simplifies to

$$u_{k+1}^{l+1} \;=\; \underset{u\,\in\,BV(\Omega)}{\arg\min} \left\{ \frac{1}{2}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u \;-\; (u_{k+\frac{1}{2}}^{l+1} \;+\; u_k^{l+1}v^l))^2}{u_k^{l+1}} \;+\; \alpha\,|u|_{BV(\Omega)} \right\}\;, \tag{5.93}$$

i.e. the second step (5.91) can be realized by a minor modification of the TV step introduced in (5.26).

The update variable v in (5.92) is an error function with respect to the optimality condition of the unregularized log-likelihood functional (5.15). In every refinement step of the Bregman iteration v^{l+1} differs from v^l by the current error in the optimality condition (5.15). Within the TV step (5.93) one observes that an iterative regularization with the Bregman distance leads to contrast enhancement. Instead of fitting to the EM solution $u_{k+\frac{1}{2}}^{l+1}$ in the weighted norm, we use a function in the fidelity term whose intensities are increased by the error function v^l. Resulting from the idea of adaptive regularization v^l is weighted by u_k^{l+1}, too. Following [124] or [182] the elements of the dual space of $BV(\Omega)$, $p^l \;=\; \frac{K^*\mathbf{1}_\Sigma}{\alpha}\,v^l \;\in\; \partial\,|u^{l+1}|_{BV(\Omega)} \;\subset\; (BV(\Omega))^*$, can be characterized as textures respectively strongly oscillating patterns. Based on this interesting interpretation, it makes sense to consider v^l as the current error function of the log-likelihood functional (5.15).

In the previous section we additionally introduced a damped variant of the FB-EM-TV algorithm. This damped strategy can be also realized very easily in each Bregman iteration step. With the abbreviation

$$q^{l+1} := \omega_k^{l+1}u_{k+\frac{1}{2}}^{l+1} \;+\; \omega_k^{l+1}u_k^{l+1}v^l \;+\; (1-\omega_k^{l+1})u_k^{l+1}\;,$$

the TV step (5.93) simply needs to be adapted to

$$u_{k+1}^{l+1} \;=\; \underset{u\,\in\,BV(\Omega)}{\arg\min} \left\{ \frac{1}{2}\int_\Omega \frac{K^*\mathbf{1}_\Sigma\,(u \;-\; q^{l+1})^2}{u_k^{l+1}} \;+\; \omega_k^{l+1}\alpha\,|u|_{BV(\Omega)} \right\}\;. \tag{5.94}$$

As usual for iterative methods the described reconstruction method by iterative regularization needs a stopping criterion. The latter should stop at an iteration offering a solution that approximates the true image as good as possible. This is necessary to prevent that too much noise arises by the inverse scale space strategy. In the case of Gaussian noise, the discrepancy principle is a reasonable stopping criterion, i.e. the procedure would stop if the residual $\left\|Ku^l - f\right\|_2$ reaches the variance of the noise. In the case of Poisson noise, however, it makes sense to stop the Bregman iteration if the Kullback-Leibler distances of Ku^l and the given data f reach the noise level. For the synthetic data the noise level δ is naturally given by the KL distance between $K\bar{u}$ and f, i.e.

$$\delta \;=\; D_{KL}(f, K\bar{u})\;,$$

where $\bar{u}$ denotes the true, noise-free image. For experimental data it is necessary to find a suitable estimate for the noise level δ from data counts. Assuming that such an estimate is available and based on the observations in this section we can use **Algorithm** 2 to solve the problem (5.89).

Algorithm 2 Bregman-EM-TV

1. **Parameters:** $f,\ \alpha > 0,\ \omega \in (0,1],\ maxBregIts > 0,\ \delta > 0,\ maxEMIts > 0,\ tol > 0$
2. **Initialization:** $k = 0,\ l = 0,\ u_0^1 = u^0 = c > 0,\ v^0 = 0$

while $(D_{KL}(f, Ku^l) \geq \delta$ **and** $l \leq maxBregIts)$ **do**

  while $(k \leq maxEMIts$ **and** $(opt_k \geq tol$ **or** $u\_opt_k \geq tol$ **or** $p\_opt_k \geq tol))$ **do** ▷ (5.30), (5.31)

    Compute $u_{k+\frac{1}{2}}^{l+1}$ via EM step analogous to (5.25) with u_k^{l+1} instead of u_k.
    Set $\omega_k^{l+1} = \omega$.
    Compute u_{k+1}^{l+1} via modified ROF model (5.94). ▷ Section 5.5
    $k \leftarrow k + 1$

  end while

  Compute update v^{l+1} via (5.92).
  Set $u_0^{l+2} = u^{l+1}$
  $l \leftarrow l + 1$;

end while
return u^l

Alternative, a dual inverse scale space strategy based on Bregman iterations can be used to obtain simultaneous contrast correction, see Subsection 3.3.3. However, both inverse scale space methods compute very similar iterates and we could not recognize a difference in the performance so far, but in the case of the dual approach we can provide error estimates and convergence rates for exact and noisy data.

Inspired by the works of [134], [34], [33], one can also think of modeling a contrast correction by iterative refinement inside the FB-EM-TV algorithm. On the basis of the two step iteration (5.25), the TV correction step can be realized by a sequence of modified ROF problems based on (5.26). More precisely, for any fixed k, u_{k+1} is

determined via a sequence of minimization problems

$$u_{k+1}^{l+1} \;=\; \underset{u\,\in\,BV(\Omega)}{\arg\min} \left\{ \frac{1}{2}\int_{\Omega} \frac{K^*\mathbf{1}_\Sigma\,(u \;-\; u_{k+\frac{1}{2}})^2}{u_k} \;+\; \alpha\,(\,|u|_{BV(\Omega)} \;-\; \langle p^l, u\rangle\,) \right\}\,, \tag{5.95}$$

$$p^l \in \partial\,|u_{k+1}^l|_{BV(\Omega)}\,,$$

with u_{k+1}^0 constant and $p^0 := 0 \in \partial\,|u_{k+1}^0|_{BV(\Omega)}$. Analogous to [134] and the derivation above, the scaling $v^l := \frac{\alpha\,u_k}{K^*\mathbf{1}_\Sigma}\,p^l$ transfers the shift term $\langle p^l, u\rangle$ to the data fidelity term in such a way that (5.95) can be rewritten similar to (5.93),

$$u_{k+1}^{l+1} \;=\; \underset{u\,\in\,BV(\Omega)}{\arg\min} \left\{ \frac{1}{2}\int_{\Omega} \frac{K^*\mathbf{1}_\Sigma\,(u \;-\; (u_{k+\frac{1}{2}} \;+\; v^l))^2}{u_k} \;+\; \alpha\,|u|_{BV(\Omega)} \right\}$$

with the following update formula

$$v^{l+1} \;=\; v^l \;+\; (u_{k+\frac{1}{2}} \;-\; u_{k+1}^{l+1})\,, \qquad v^0 \;=\; 0\,,$$

which seems rather related to an additive than a multiplicative setting and hence less promising for our setting. Computational experiments indeed confirm that the inner correction leads to worse reconstructions than the outer one.

5.6.2 Dual-Bregman-EM-TV

In Section 3.3.3 we presented a dual inverse scale space method in terms of an iterative Bregman regularization technique for general, convex data fidelities and regularization terms. This strategy based on a dual representation of the initial variational problem (3.3.1). A bidual formulation of the dual inverse scale space strategy offers a simple interpretation in terms of a familiar (primal) problem (3.11). In the special case of Poisson noise modeling and TV regularization, this reads as follows

$$u^{l+1} = \underset{\substack{u\in BV(\Omega)\\ u\geq 0}}{\arg\min} \left\{ \int_\Sigma (Ku + r^l - f\log(Ku + r^l))\;d\mu \;+\; \alpha|u|_{BV} \right\}\,, \tag{5.96}$$

with the update of the residual function r^l (see (3.12))

$$r^{l+1} = r^l + Ku^{l+1} - f\,, \qquad \text{with } r^0 = 0\,. \tag{5.97}$$

Now we can compare the variational problem in (5.96) with the Kullback-Leibler TV problem in (5.18) concerning an inverse problem with a background model b, i.e.

$Ku + b = f$. This reveals that the noise function r^l serves as a dynamically updated background model instead of a time-constant background b. Shifting the argument of the data fidelity with r^l in that appropriate way, leads to the expected contrast enhancing behavior in each time step $l \to l+1$.

Although the minimization problem (5.96) for a specific l can intuitively be implemented similarly to the splitting strategy of EM-TV,

$$\begin{cases} u_{k+\frac{1}{2}} &= u_k \, K^* \left(\dfrac{f}{Ku_k + r^l} \right) & \text{(EM step)} \\ u_{k+1} &= u_{k+\frac{1}{2}} - \alpha \; u_k \; p_{k+1} & \text{(TV step)} \end{cases},$$

we need to be aware of division-by-zero problems in the EM step. For the dual inverse scale method, we can overcome this problem by a partially explicit approximation. For this sake we rewrite the optimality condition in the following way, using $K^*1 = 1$,

$$\begin{aligned} & 1 - K^* \left(\frac{f}{Ku + r} \right) + \alpha p = 0 \\ \Longrightarrow \; & 1 - \frac{f}{Ku + r} + \alpha (K^*)^{-1} p = 0 \\ \Longrightarrow \; & Ku + r - f + \alpha (K^*)^{-1} p (Ku + r) = 0 \\ \Longrightarrow \; & K^* \left(1 - \frac{f - r - \alpha r (K^*)^{-1} p}{Ku} \right) + \alpha p = 0 \\ \Longrightarrow \; & 1 - K^* \left(\frac{f - r(1 + \alpha (K^*)^{-1} p)}{Ku} \right) + \alpha p = 0 \, . \end{aligned}$$

Now we use an approximation of the first term including the subgradient p from the last Bregman step and obtain

$$u_{k+1} = u_k \, K^* \left(\frac{f - r^l (1 + q^l)}{K u_k} \right) \; - \; \alpha \, u_k \, p_{k+1} \, ,$$

with $q^l = \alpha (K^*)^{-1} p^l$. Note that q^l does not need to be computed by inverting K^*, but can be obtained from the update formula

$$q^{l+1} = \frac{f - r^l (1 + q^l)}{K u^{l+1}} - 1. \tag{5.98}$$

Based on these observations we can use **Algorithm** 3 to realize the dual inverse scale space method in (5.96) and (5.97).

Algorithm 3 Dual Bregman-EM-TV

1. *Initialization* : $u_0^0 := c > 0,\ r^0 = 0,\ q^0 = 0$
2. *For* $l = 0, 1, 2, \ldots$:
   Compute u^{l+1} *via iteration scheme* 2. *in*
   Algorithm 1, *but with* $f - r^l(1 + q^l)$
   instead of f *in EM steps* .
   Then update q^{l+1} *via* (5.98) .
   Then update the residual r^{l+1} *via* (5.97) .

6 3D Reconstruction in Nanoscopy and PET

6.1 Applications

In this section we will illustrate the proposed algorithms and we will test the theoretical results by 2D and 3D synthetic and experimental results in high resolution fluorescence microscopy and positron emission tomography.

6.1.1 Optical Nanoscopy

In recent years revolutionary imaging techniques have been developed in light microscopy with enormous importance for biology, material sciences, and medicine. The technology of light microscopy has been considered to be exhausted for a couple of decades, since the resolution is basically limited by Abbe's law for diffraction of light. By developing stimulated emission depletion (STED)- and 4Pi-microscopy [97] now resolutions are achieved that are way beyond this diffraction barrier [112, 96]. STED-microscopy [191] takes an interesting laser sampling approach, which in principle would even allow molecular resolutions. Fluorescent dyes are stimulated by a small laser spot and are directly quenched by an additional interfering laser spot. Since this depletion spot vanishes at one very small point in the middle, fluorescence of the simulating spot is only detected at this tiny position. Hence, data with previously unknown resolution can be measured. However, by reaching the diffraction limit of light, measurements suffer from blurring effects and in addition suffer from Poisson noise due to laser sampling.

In the case of optical nanoscopy the linear, compact operator K describes a convolution operator with a kernel $k \in C(\Omega \subset \mathbb{R}^d)$

$$(Ku)(x) \;=\; (k * u)(x) \;:=\; \int_\Omega k(x-y)\, u(y) dy \;. \tag{6.1}$$

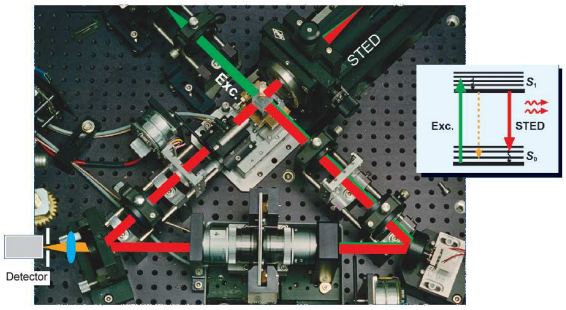


Figure 6.1: Optical system of a STED microscopy, MPI Göttingen

The kernel is often referred to as the point spread function (PSF), whose Fourier transform is called object transfer function. From a computational point of view, it is important to say that the convolution operator in the proposed algorithms can be computed efficiently by FFT following the Fourier convolution theorem,

$$k * u \;=\; \mathcal{F}^{-1}\left(\mathcal{F}(k) \cdot \mathcal{F}(u)\right) \;.$$

6.1.2 Positron Emission Tomography (PET)

Positron emission tomography (PET) is a biomedical imaging technique, which enables to visualize biochemical and physiological processes, such as glucose metabolism, blood flow, or receptor concentrations, see e.g. [189], [178], [10]. This modality is mainly applied in nuclear medicine and can be used for instance to detect tumors, locate areas of the heart affected by coronary artery disease and identify brain regions influenced by drugs. Therefore, PET is categorized as a functional imaging method and differs from methods such as X-ray computed tomography (CT) that depicts priori anatomy structures. The data acquisition in PET is based on weak radioactively marked pharmaceuticals, so-called tracers, which are injected into the blood circulation, and bindings dependent on the choice of the tracer to the molecules to be studied. Used markers are suitable radio-isotopes which decay by emitting a positron, which annihilates almost immediately with an electron. The resulting emission of two photons is detected by detector rings surrounding the patient. Due to the radioactive decay, measured data can be modeled as an inhomogeneous Poisson process with a mean given by the X-ray transform of the spatial tracer distribution. The X-ray transform maps a function on $\mathbb{R}^d$ into the set of its line integrals. More precisely, if $\theta \in S^{d-1}$ and $x \in \Omega$,

$$(Ku)(\theta, x) \;=\; \int_{\mathbb{R}} u(x \,+\, t\,\theta)\; dt \;, \qquad x \,+\, t\,\theta \;\subseteq\; \Omega \;,$$

is the integral of u over the straight line through x with direction θ. Up to notation, in the 2-dimensional case the X-ray transform is equivalent to the more popular Radon transform.

6.2 Results

6.2.1 2D Deconvolution in Optical Nanoscopy

Synthetic Results (2D):
To get an impression of images suffering from blurring effects and Poisson noise, we refer to Figure 6.3. Exemplary, you can see a synthetic data set in 6.3(a) regarding a special 4Pi convolution kernel, see Figure 6.3(b), (c). Compared to standard convolution kernels, e.g. of Gaussian-type, the latter bears an additional challenge since it varies considerably in structure. This leads to side lobes in the object structure of the measured data 6.3(d), (e). In practice, this type of convolution can be found for instance in 4Pi microscopy [97], since two laser beams interfere in the focus. Under certain circumstances, convolution kernels can also be locally varying, such that blind deconvolution strategies are in need. In this case, an additional unknown phase variable ϕ influences the structure of the underlying convolution kernel, see Figure 6.2. For blind deconvolution strategies cf. [Stück and Hohage].

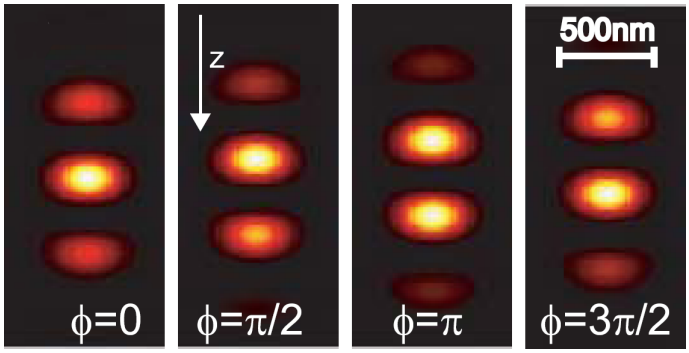


Figure 6.2: Convolution kernels in 4Pi microscopy for different phases

In this section we assume a 4Pi convolution kernel in the form

$$h(x,y) \;\sim\; cos^4\left(\left(\frac{2\Pi}{\lambda}\right)y\right) e^{-\left(\frac{x}{\sigma_x}\right)^2-\left(\frac{y}{\sigma_y}\right)^2}, \tag{6.2}$$

with the standard deviations σ_x and σ_y, and where λ denotes the refractive index characterizing the doubling properties.

At the beginning we will illustrate our proposed techniques for the introduced synthetic object (cf. Figure 6.3(a)), blurred with the introduced 4Pi convolution kernel (6.2). The given data additionally suffer from Poisson noise, see Figure 6.3(e).

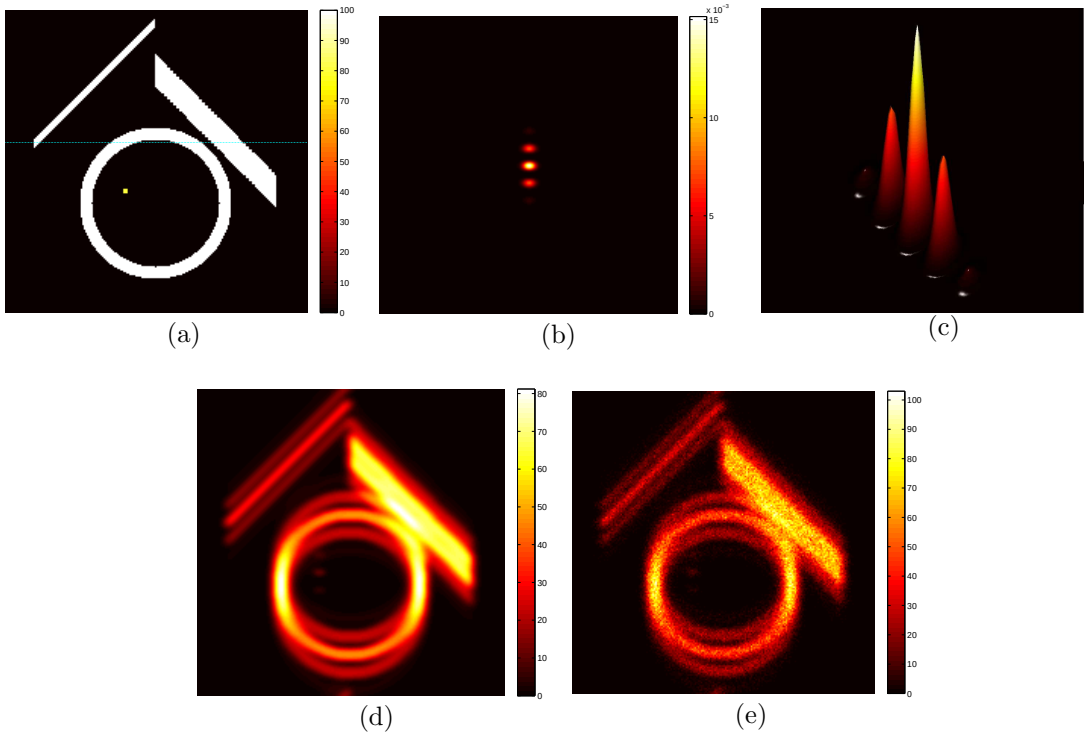


Figure 6.3: **2D synthetic data with 4Pi blurring and Poisson noise**: (a) Desired image (ground truth), resolution $|\Omega| = 200^2$; (b) Assumed 4Pi microscopy PSF with parameters $\lambda = 0.12, \sigma_x = 0.02, \sigma_y = 0.07$; (c) 3d visualization of the 4Pi PSF; (d) Convolved using 4Pi PSF, and (e) Given data, convolved with 4Pi PSF and in addition suffering from Poisson noise.

In Figure 6.4 we present EM reconstructions following algorithm (5.17) and we present the corresponding multiplicative residual images $\frac{u_{EM}}{uexact}$ for different numbers of iterations. Early stopping leads to natural regularization and undesired side lobes remain in the EM reconstruction. A higher number of iterations leads to the well known, undesired checkerboard effect. In Figure 6.4(g), (h) we display the KL-distances in the data and image domain. The former decreases and reaches the exact reference distance $H(f, Ku_{EM})$ approximately after 50 iterations, whereas the latter starts to increase again after initial decrease. In Figure 6.5 and Figure 6.6 we illustrate different FB-EMTV reconstructions following (5.25) and statistical

investigations concerning optimality. In comparison to EM, FB-EMTV deconvolves the given data without remaining side lobes and reduces noise and oscillations very well, see 6.5(a),(b),(c) and the slices in (g). For $\alpha = 0.04$ the reconstruction is slightly under-smoothed, whereas for $\alpha = 0.2$ the computed image is over-smoothed. A visually reasonable reconstruction can be obtained for $\alpha = 0.02$. The corresponding plots of the optimality condition, the functional values and the Kullback-Leibler measurements show an expected decreasing behavior. In 6.6(h) we visualize different values of the regularization parameter α versus the KL-distances due to FB-EMTV reconstructions with α. Although the underlying object can be reconstructed quite well with FB-EMTV, we can observe a natural loss of contrast as mentioned in 5.6.1. Parts of the test object cannot be separated sufficiently. One possibility to overcome this problem is the use of inverse scale space methods.

In Section 5.3.2 and 5.4.4 we found out that a damping strategy in FB-EMTV is in general necessary to guarantee convergence. For deblurring our testobject the necessity of damping can be observed for higher values of the regularization parameter α. In Figure 6.7 we compare functional values for different damping parameters ω in the case of $\alpha = 10$. Without damping ($\omega = 1$) we obtain oscillations in Figure 6.7(a), whereas for $\omega = 0.07$ a monotone descent in the objective functional can be achieved 6.7(c). However, due to our numerical tests in many cases damping is not needed to achieve a convergent behavior of the FB-EMTV algorithm.

The Bregman-EMTV strategy proposed in Section 5.6.1 improves the reconstruction considerably due to simultaneous contrast enhancement, see Figure ... . This aspect is underlined by the values of the KL-distance for the different reconstructions.

Experimental Results in Optical Nanoscopy (2D):
In Figure 6.9 we illustrate our techniques by reconstructing Syntaxin [191], a membrane integrated protein participating in exocytosis. Here, the contrast enhancing property of Bregman-EM-TV is observable as well, compared to EM-TV. It is possible to preserve fine structures in the image.

6.2.2 2D PET Reconstruction in Nuclear Medicine

In figure 6.10, we illustrate our techniques by evaluation of cardiac $H_2\,^{15}O$ measurements obtained from PET. This tracer is used for the quantification of myocardial blood flow [162]. This quantification needs a segmentation of myocardial tissue, left and right ventricle [162], [19], which is usually extremely difficult to realize due to very low SNR of $H_2\,^{15}O$ measurements. In order to obtain the tracer intensity in the

right and left ventricle we take a fixed 2D layer in two different time frames (25 and 45 seconds after tracer injection in the blood circulation).

6.2.3 2D Primal and dual Bregman in Optical Nanoscopy

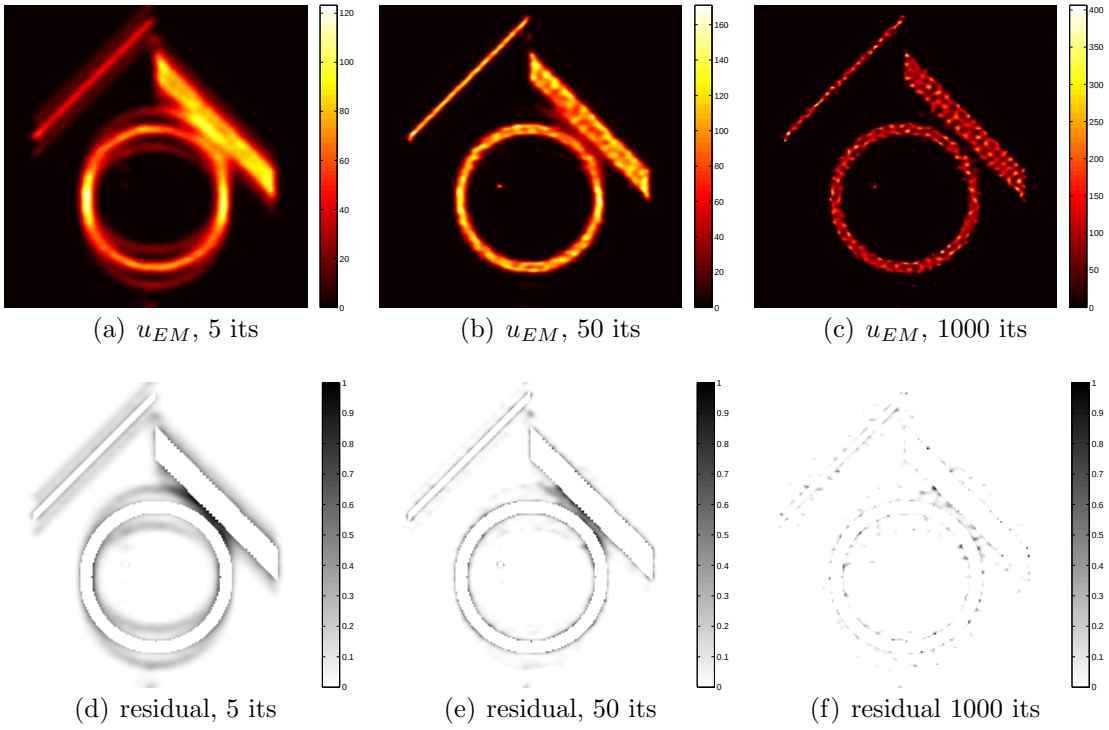
In this section we present the performance of the proposed techniques by reconstructing synthetic and experimental data. Figure 6.11 illustrates our techniques at a simple synthetic object by applying a 4Pi convolution and adding Poisson noise. With EM-TV (see 6.11(c) and 6.11(g)) we get rid of noise and oscillations, but we are not able to separate the objects sufficiently. Using Bregman-EM-TV a considerable improvement resulting from contrast enhancement can be achieved. This aspect is underlined by the values of the KL-distance for the different reconstructions. In Figure 6.13 we compare the primal and dual inverse scale space strategy using the same synthetic object, but in this case with a Gaussian convolution kernel. As expected, both inverse scale space strategies compute very similar iterates and we can observe a decrease of the Kullback-Leibler distance between u and $\tilde{u}$ until the noise level is reached. Taking a closer look at the distance measurements reveals a slightly better decrease in the case of Dual-Bregman-EM-TV.

Figure 6.12, (a)-(c) demonstrate the protein Bruchpilot [110] and its EM-TV and Bregman-EM-TV reconstruction. Particularly, the latter delivers well separated object segments and a high contrast level. It is possible to preserve fine structures in the image. Due to the Fourier convolution theorem, the convolution operator (6.1) can be computed efficiently via FFT

$$k * u \;=\; \mathcal{F}^{-1}\left(\mathcal{F}(k) \cdot \mathcal{F}(u)\right) \;,$$

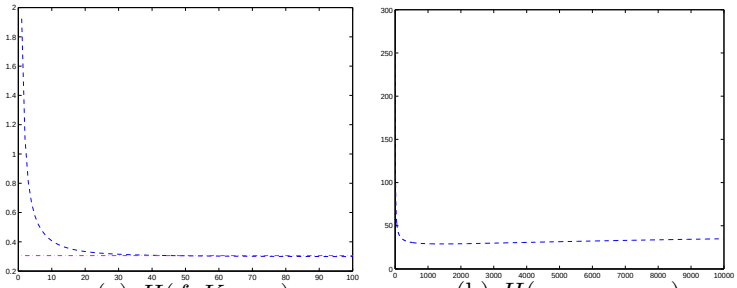
such that the EM steps in the proposed algorithms can be performed quickly.

6.2.4 3D Deconvolution in Optical Nanoscopy



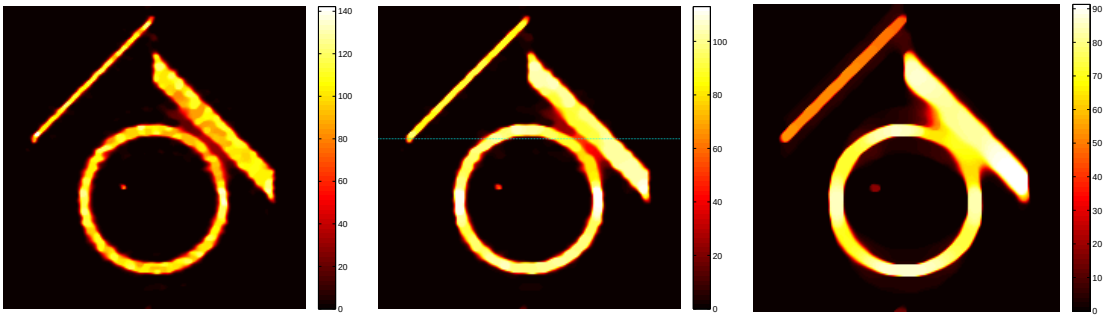
(a) u_{EM}, 5 its (b) u_{EM}, 50 its (c) u_{EM}, 1000 its

(d) residual, 5 its (e) residual, 50 its (f) residual 1000 its

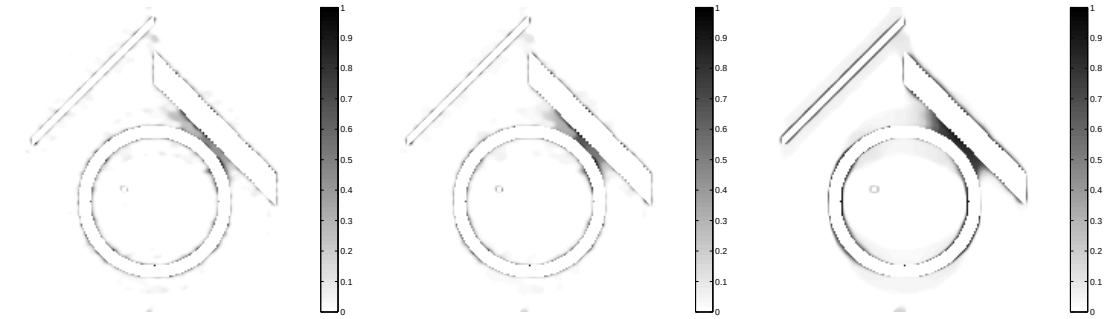


(g) $H(f, Ku_{EM})$ (h) $H(u_{EM}, u_{exact})$

Figure 6.4: **EM reconstructions**: (a)-(c) Reconstructions following algorithm (5.17) stopped at different iterations numbers; (d)-(f) Corresponding residuals $\frac{u_{EM}}{u_{exact}}$, scaled to $[0, 1]$; (g) Data KL-distance $H(f, Ku_{EM})$ for 100 iterations (blue dashed line) and reference value $H(f, Ku_{exact})$ (magenta dash-dot line), and (h) Image KL-distance $H(u_{EM}, u_{exact})$ for 10000 iterations.

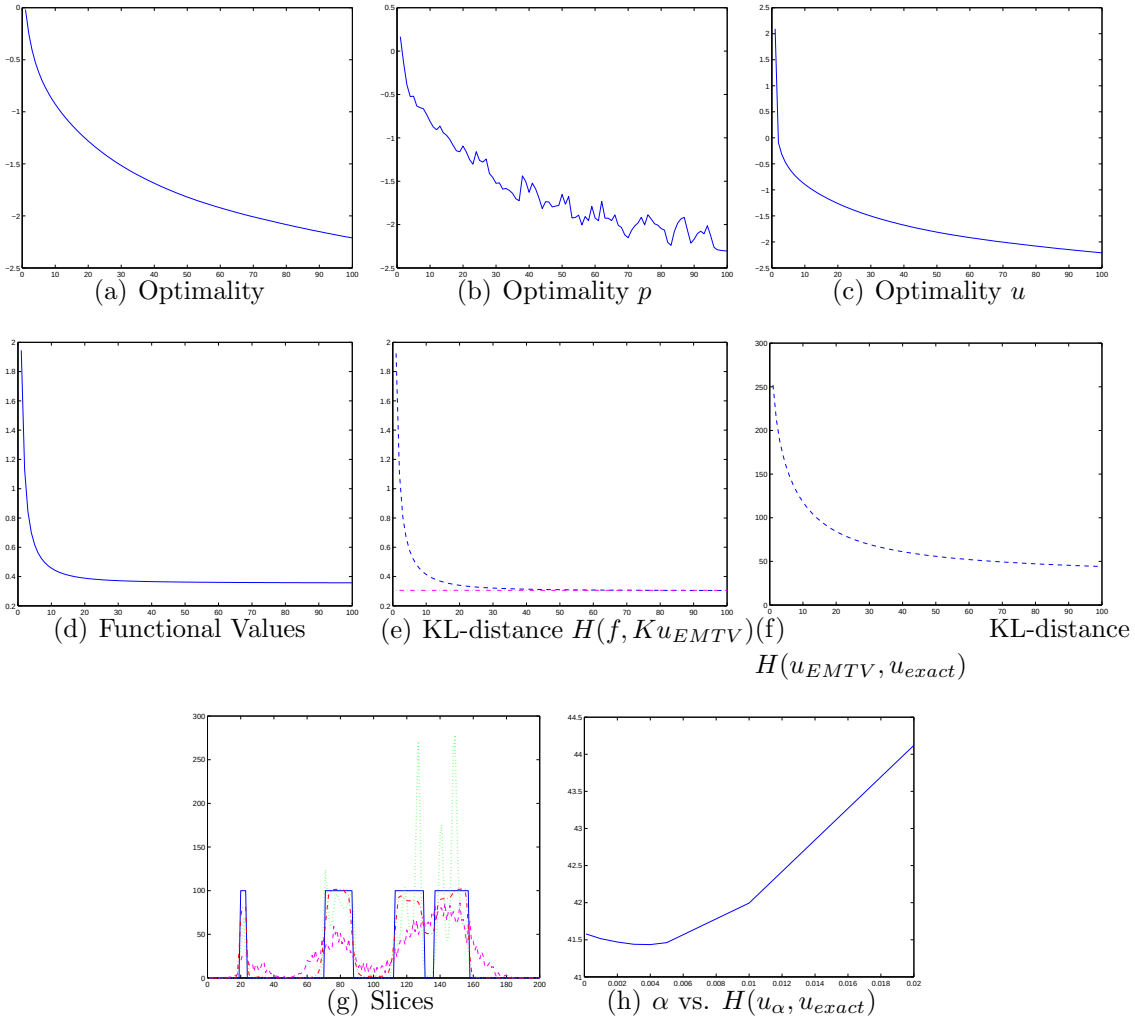


(a) $\alpha = 0.04$ (b) $\alpha = 0.02$ (c) $\alpha = 0.2$



(d) Residual $\alpha = 0.004$ (e) Residual $\alpha = 0.02$ (f) Residual $\alpha = 0.2$

Figure 6.5: **FB-EMTV reconstructions**: (a)-(c) Reconstructions following algorithm (5.25) for different regularization parameters α, and (d)-(f) Corresponding residuals $\frac{u_{EMTV}}{u_{exact}}$, scaled to $[0, 1]$.

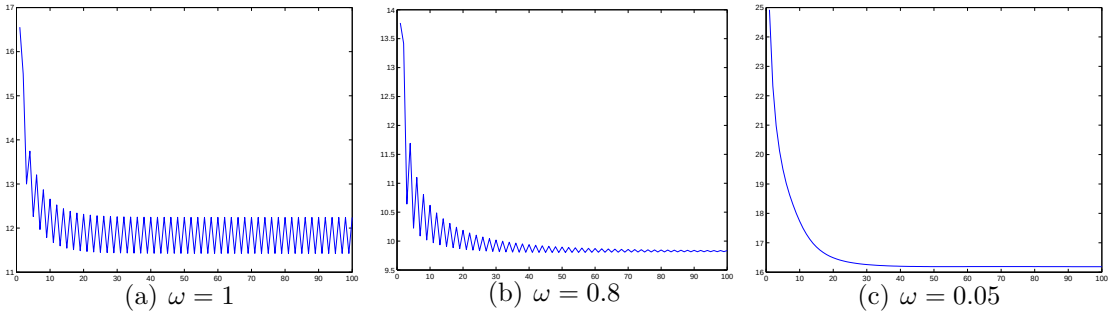


(a) Optimality (b) Optimality p (c) Optimality u

(d) Functional Values (e) KL-distance $H(f, Ku_{EMTV})$ (f) KL-distance $H(u_{EMTV}, u_{exact})$

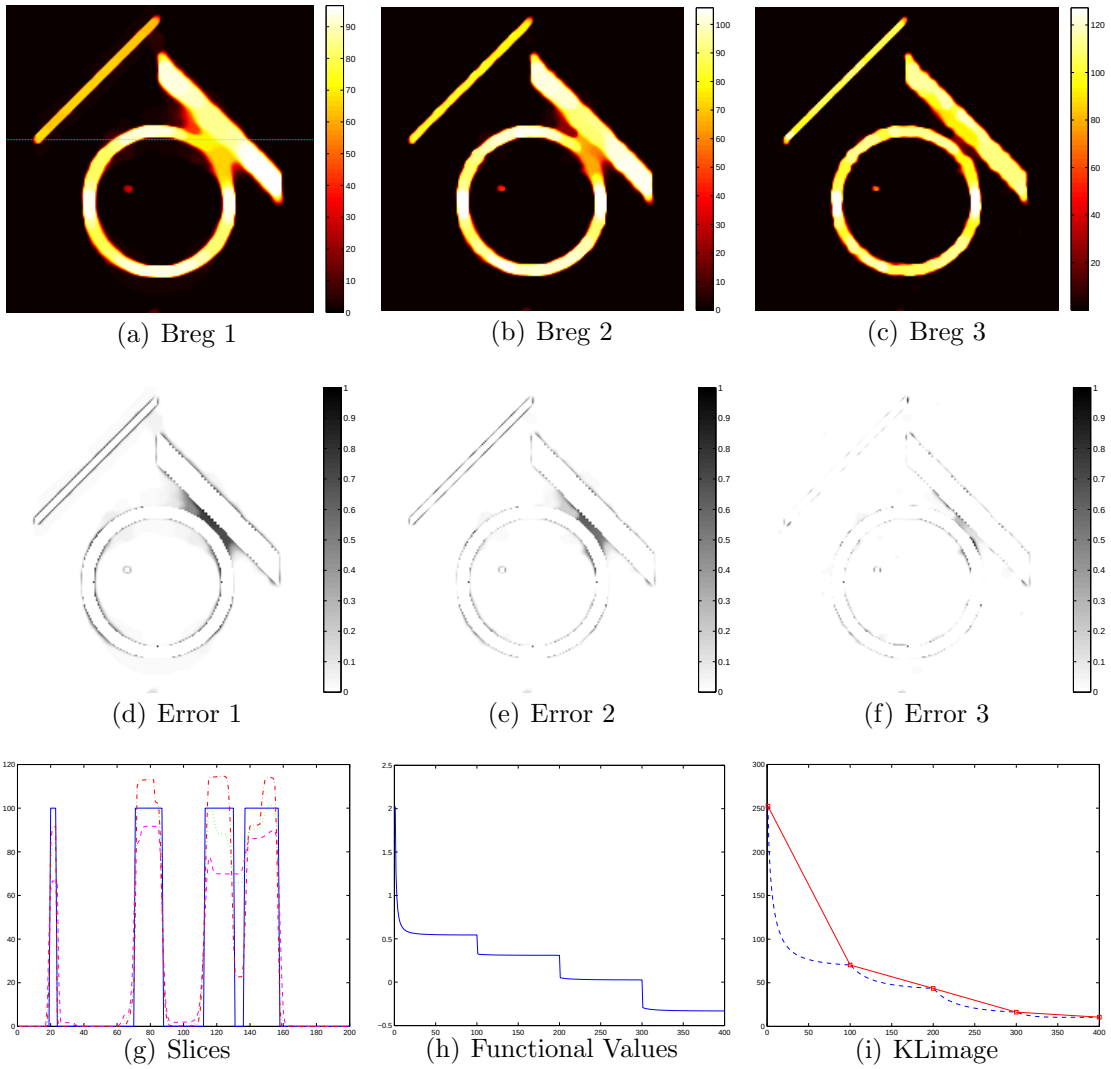
(g) Slices (h) α vs. $H(u_\alpha, u_{exact})$

Figure 6.6: **EM-TV Reconstructions, Statistics**



(a) $\omega = 1$ (b) $\omega = 0.8$ (c) $\omega = 0.05$

Figure 6.7: **Damping in FB-EMTV**: iterations vs. values of the objective functional for different values of the damping parameter ω for $\alpha = 10$.

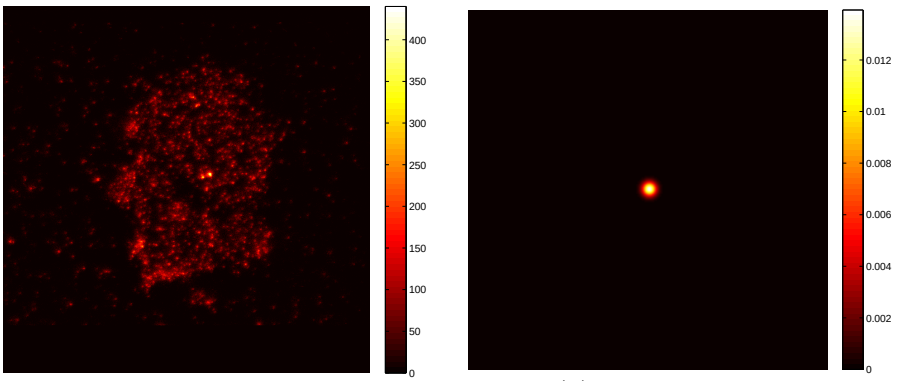


(a) Breg 1 (b) Breg 2 (c) Breg 3

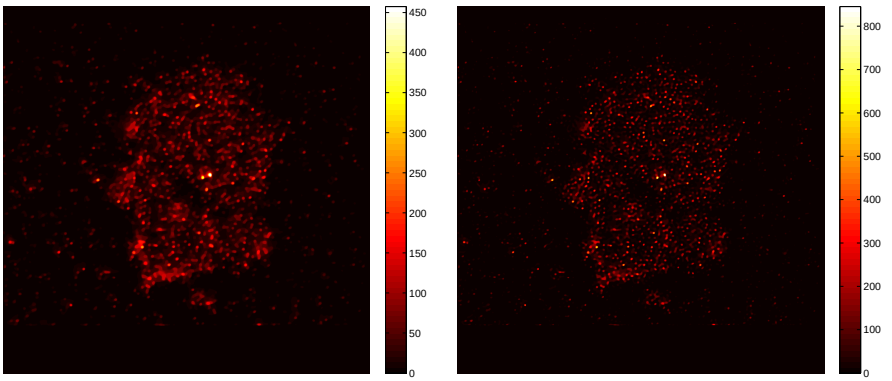
(d) Error 1 (e) Error 2 (f) Error 3

(g) Slices (h) Functional Values (i) KLimage

Figure 6.8: **Bregman-EM-TV**:



(a) Syntaxin (b) psf 3d visualization



(c) EM-TV (d) Bregman-EM-TV

Figure 6.9: 2D experimental data: (a) Protein Syntaxin in cell membrane, fixed mammalian (PC12) cell; image size: 1000 x 1000; (b) 3D visualization of the used point spread function; (c) EM-TV reconstruction; and (d) 3rd iterate u^3 of the Bregman-EM-TV algorithm.

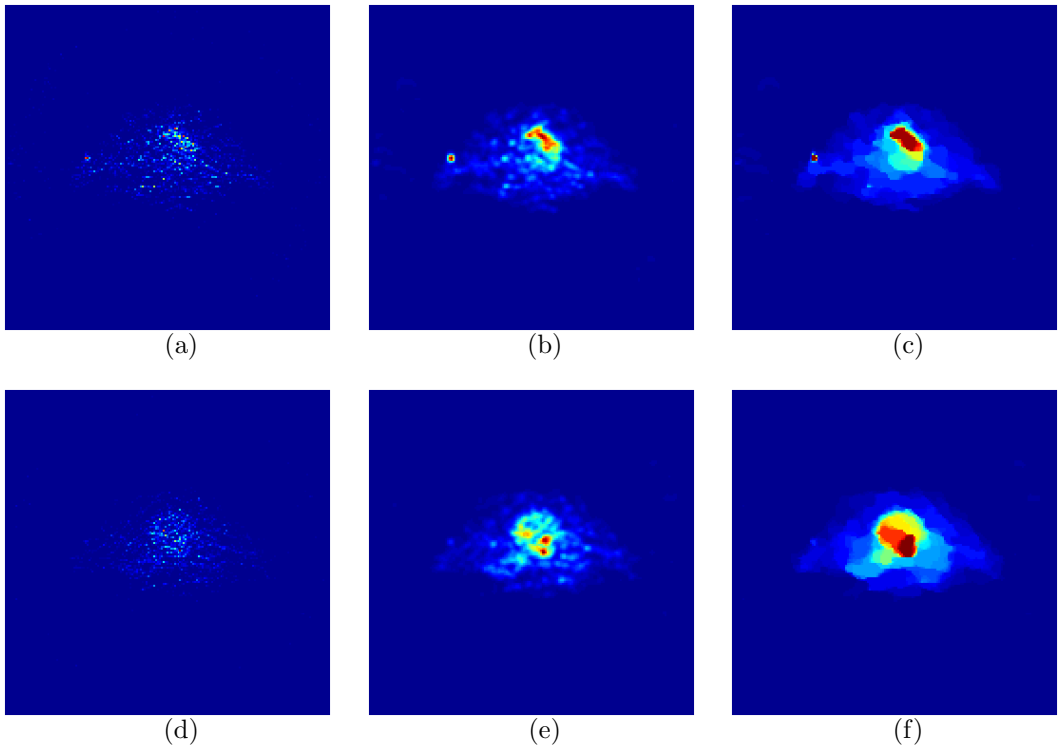


Figure 6.10: Cardiac $H_2{}^{15}O$ PET measurements: results of different reconstruction methods in two different time frames.

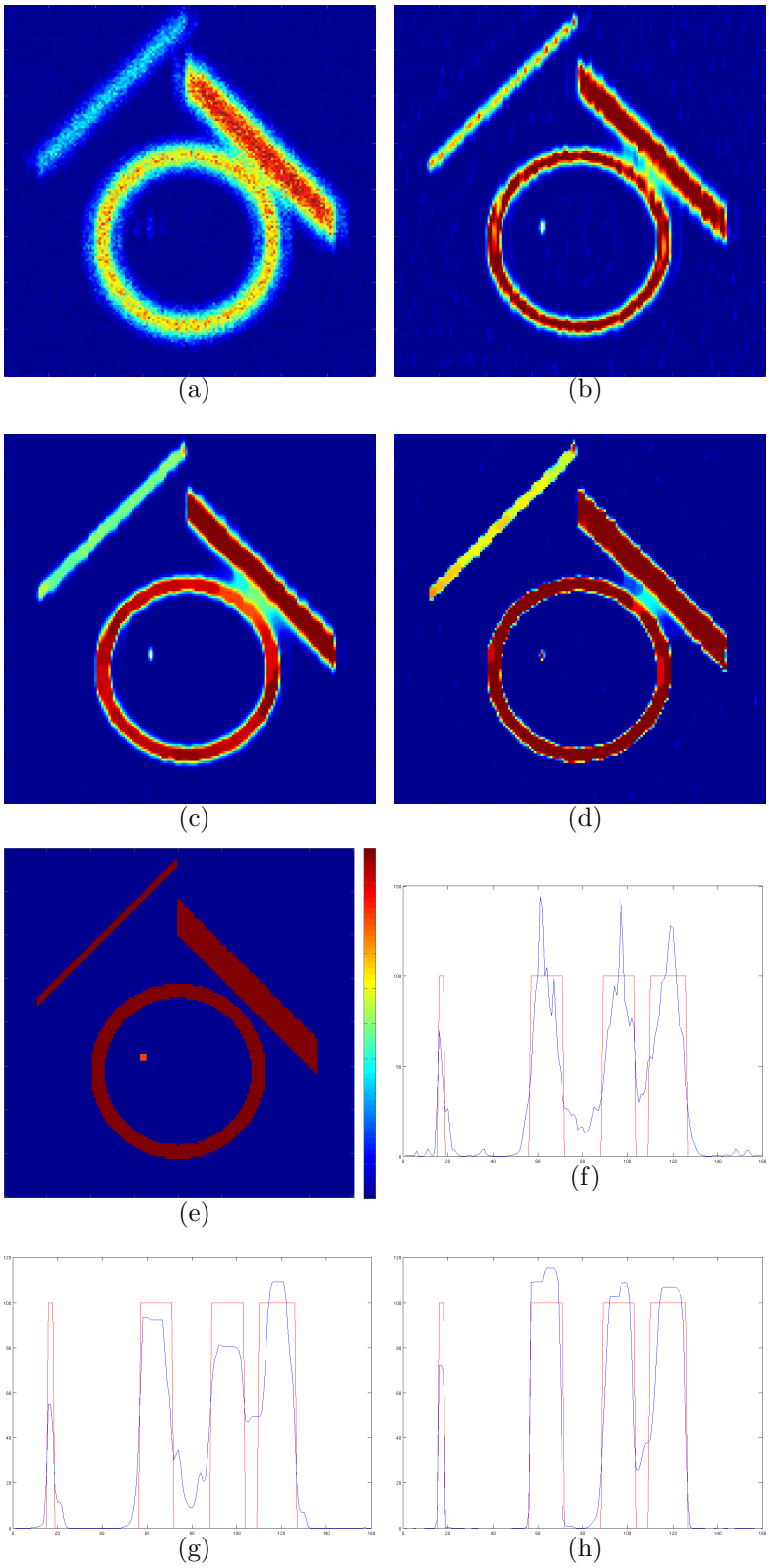
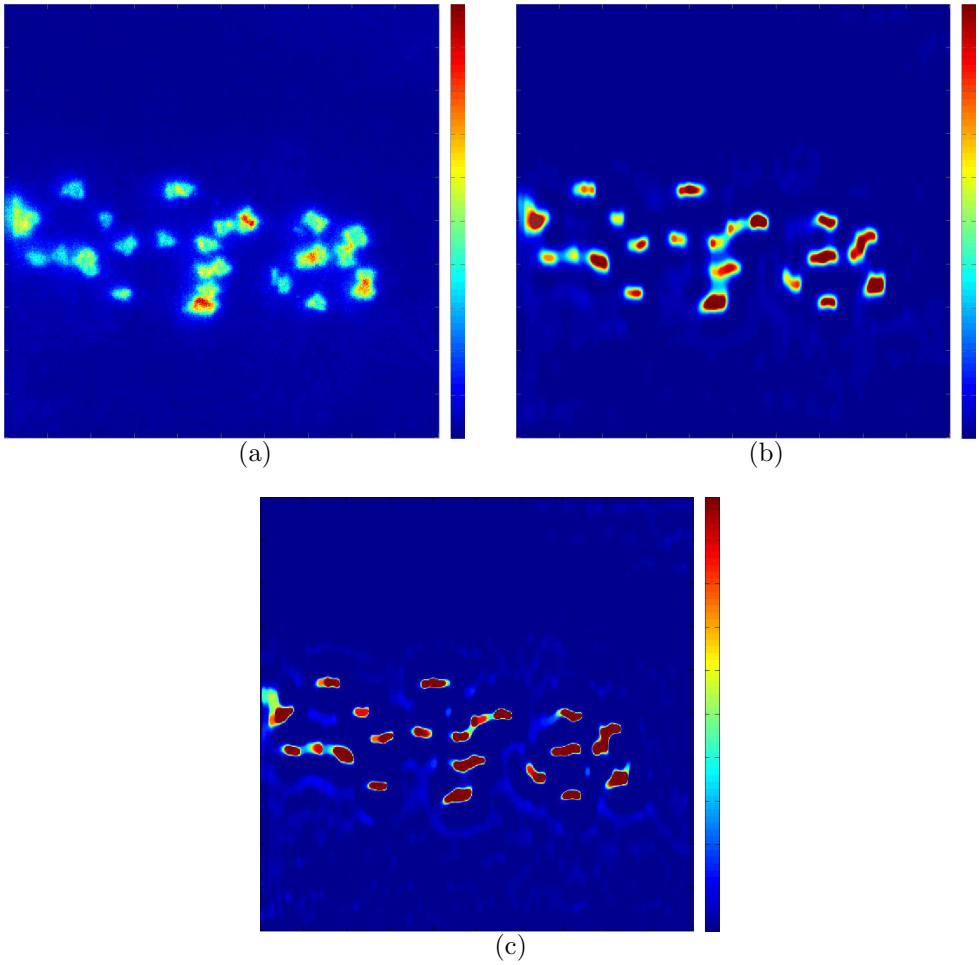


Figure 6.11: Synthetic data: (a) raw data using 4Pi PSF; (b) EM reconstruction, 20 its, KL-distance: 3.20; (c) EM-TV, $\alpha = 0.04$, KL-distance: 2.43; (d) Bregman-EM-TV, $\alpha = 0.1$, after 4 updates, KL-distance: 1.43; (e) true image; (f)-(h) horizontal slices EM, EM-TV and Bregman-EM-TV compared to true image slice.



(a) (b) (c)

Figure 6.12: Experimental data: (a) Protein Bruchpilot in active zones of neuromuscular synapses in larval Drosophila; (b) EM-TV; (c) Bregman-EM-TV;

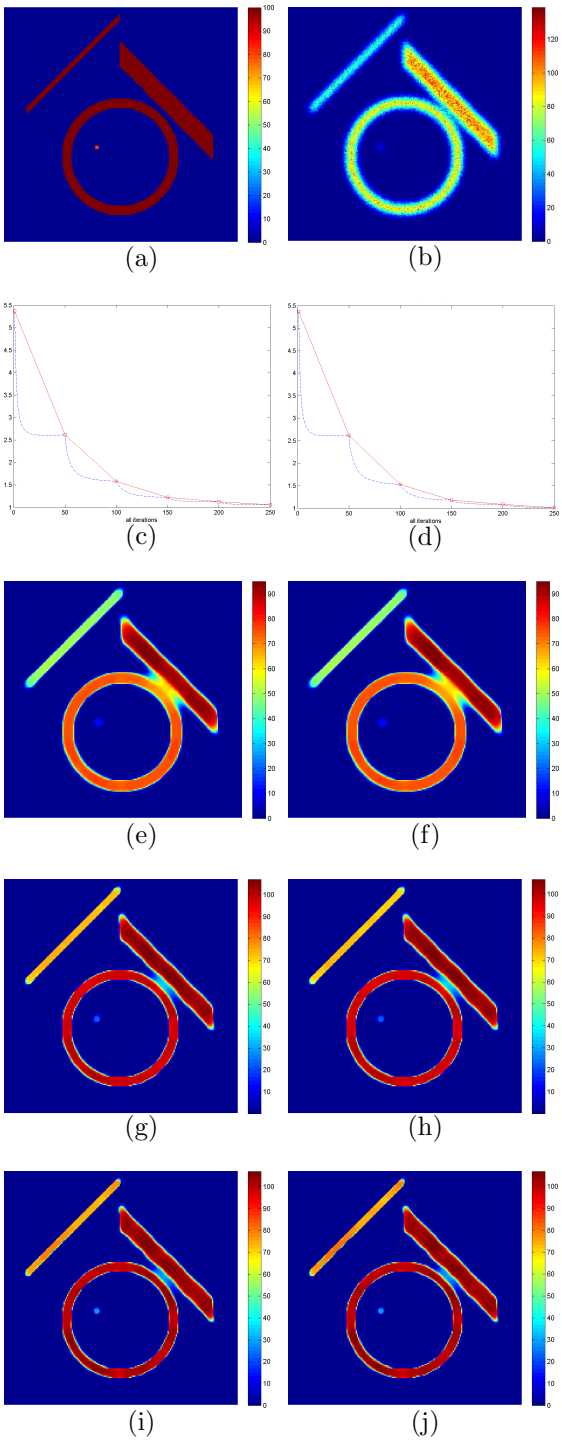
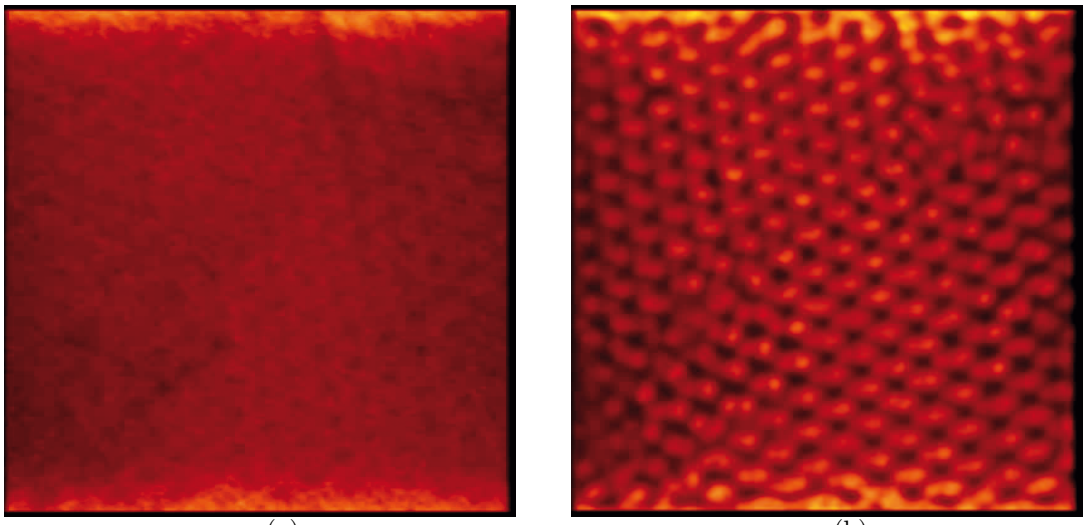


Figure 6.13: Synthetic data: Comparison of primal and dual inverse scale space methods; (a): true image; (b): raw data f using Gaussian PSF; (c)-(d): KL-distance between u and $\tilde{u}$ for Bregman-EM-TV resp. Dual-Bregman-EM-TV, blue line: distance at all 250 iterations, red marker: distance at every Bregman step (intervals of 50 interior iterations)(e),(g),(i): iterates u^1,u^3 and u^5 of Bregman-EM-TV; (f),(h),(j): iterates u^1,u^3 and u^5 of Dual-Bregman-EM-TV.



(a) (b)

Figure 6.14: 3D experimental data

7 4D Imaging

7.1 Introduction

This chapter serves as a "bridge" between 3D static image reconstruction and 4D time dependent imaging. In the previous chapters we studied models and algorithms for 3D static image reconstruction in the applications tomography and high resolution fluorescence microscopy. In the second part of this thesis we will extend 3D image reconstruction to 4D image sequence reconstruction. Instead of a static inverse problem, $Ku = f$, we now have to solve inverse problems in space and time, i.e.

$$\tilde{K}(\rho(x,t)) \;=\; f(x,t), \qquad (x,t) \in \Omega \times [0,T] \subset \mathbb{R}^3 \times \mathbb{R} \;.$$

For fixed time steps $t = t_0$, intuitively, one may take into account the application of static reconstruction techniques similarly to the first part of the thesis. However, since we talk about time dependent inverse problems, we have to consider that "life is always in motion". In particular, in our applications in 4D medical imaging and video fluorescence microscopy we can observe natural motion effects. For example in positron emission tomography we have to consider natural patient motion like breathing or beating of the heart. To be more precise, in clinical studies it has been found that the breathing motion of the diaphragm can range from 4 mm to 38 mm, cf. [179].

Similarly, in dynamic fluorescence microscopy (e.g. dynamic STED) we can think of reconstructing videos in live cell imaging. Here we have motion effects for instance due to the migration of single cells or proteins. In optical nanoscopy time dependent reconstruction algorithms and tracking techniques are becoming more and more important. For example, recently, biophysicist achieved to observer and study living cells in nano-scale, cf. [95, 115].

For the reconstruction of 4D data in medical imaging or video microscopy standard methods as simple generalizations of 3D reconstruction algorithms can be applied, but usually they do not incorporate time dependent motion models. However, in dynamic tomography, e.g. in positron emission tomography (PET), motion is a well-known source of degradation of reconstruction results. Hence reconstruction models without incorporating motion models can cause significant blurring effects in resulting image sequences, especially at object boundaries, cf. Figure 7.1 (artifacts at the cardiac wall in the bottom image). Here we can see undesired doubling effects.

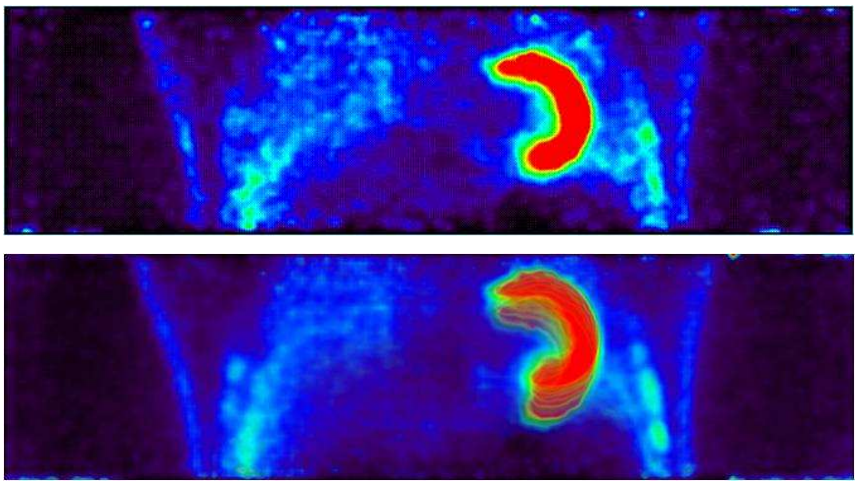
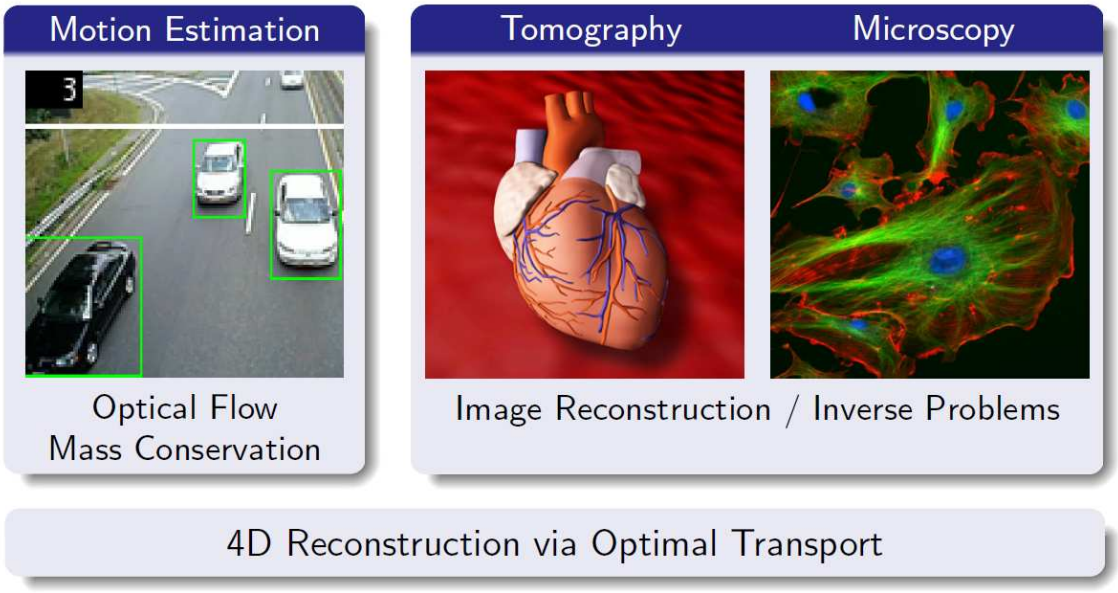


Figure 7.1: Comparison of a PET sequence reconstruction with and without motion modeling; Blurring effects due to motion artifacts in the bottom image. The data is from the European Institute for Molecular Imaging (EIMI).

The blurring effects on the PET images usually are proportional to the magnitude of the motion. In the literature it is well-known that these motion artifacts can cause significant errors in a later quantification. For example motion artifacts can cause a wrong staging of tumors, cf. [135, 66], or they can cause incorrect uptake values, cf. [132], and it can also happen that small tumors may remain undetected [137]. Therefore, the combination of parameter estimation models or motion models and reconstruction techniques is an interesting recently growing research area, cf. [123, 144, 80, 54, 147, 181].

As a consequence, our first goal for the rest of this thesis is to study and characterize various models for motion estimation and optimal transport. Our final goal is to combine the concepts of motion estimation and optimal transport with the reconstruction ideas of the first part of the thesis to build a joint 4D model for simultaneous image reconstruction, total variation regularization and optimal transport (including mass conservation), see Figure 7.1. In the following sections we will



give a short introduction on motion models and optimal transport and touch on the new joint 4D model for simultaneous motion estimation and image reconstruction.

7.2 Optical Flow and Tracking

In computer vision optical flow estimation deals with the computation of visual motion information in image sequences. For a given video $\rho(x,t)$ a common assumption for estimating the optical flow is brightness constancy,

$$\rho(x+\mathbf{v},t+\delta t) \stackrel{!}{=} \rho(x,t) \ ,$$

i.e. following the desired velocity field $\mathbf{v}$, the pixel intensity should not change. By using a Taylor expansion this assumption implies the following PDE, the so-called optical flow constraint (OFC),

$$\partial_t \rho + \mathbf{v} \ \nabla_\Omega \rho = 0 \ .$$

Hence, a standard model for optical flow estimation reads as follows,

$$\min_{\mathbf{v}} \ \|\partial_t \rho + \mathbf{v} \ \nabla_\Omega \rho\|_2^2 \ + \ \alpha \, J(u) \ ,$$

where the first term, the data fidelity term, penalizes deviations from the OFC and where $J(u)$ denotes a regularization term penalizing deviations from a certain smoothness of the flow field. In the following chapter 8 we will discuss different

motion models based on optical flow further in detail. We will discuss different data fidelities and different smoothing terms for the flow field. In particular, we will combine optical flow techniques with total variation regularization and splitting techniques, and we will present results in high resolution computed tomography (CT) and tracking applications.

For dynamic tomography data of the heart we basically have two types of motion. On the one hand motion due to the respiratory displacement and on the other hand motion due to the cardiac contraction. We will point out that the optical flow concept only takes into account incompressible flows, whereas mass conservation is more general and also takes into account compressible flows which will be useful e.g. for cardiac contraction. In other words, mass conservation allows significant density changes due to time, whereas the mass is preserved. Hence, we pass over to continuity equations and optimal transport.

7.3 Optimal Transport and Mass Conservation

In Chapter 9 we will give an introduction to basic concepts of continuum mechanics. Based on this, we will study mass conservation and optimal transport further in detail. Particularly with regard to our joint 4D reconstruction model we will concentrate the following model for optimal transport with mass conservation

$$\inf_{\rho,\mathbf{v}} \; T \int_0^T \int_\Omega \rho(x,t) \; |\mathbf{v}(x,t)|^2 \, dx \, dt$$

$$\begin{aligned} \text{subject to} \qquad \partial_t \rho + \nabla_\Omega \cdot (\rho \, \mathbf{v}) &= 0 \qquad \text{(mass conservation)} \\ \rho(\cdot, 0) &= \rho_0 \\ \rho(\cdot, T) &= \rho_T \; , \end{aligned}$$

which was introduced by Benamou and Brenier in [17, 18]. The main idea of this constraint optimization problem is to find an optimal transport "plan" $\mathbf{v}$, to move a given start density ρ_0 to a given end density ρ_T (the original problem of Monge [128] was to move a heap of sand into a hole of the same size). The continuity equation is the transport equation and the integral represents the transport cost.

Models for optical flow computation and models for mass conservation respectively optimal transport are directly related to standard models for image registration, see e.g. [126, 127, 91, 90, 92, 159]. One main difference of standard registration techniques and the flow techniques presented above lies in the use of different coordinates.

In image registration one is usually interested in finding a grid transformation, which is related to the so-called Euler coordinates, whereas in the methods above one is interested in the flow that "drives" the intensities, which is related to the so-called material or Lagrangian coordinates.

7.4 Full Joint 4D Model

At the end of this thesis in Chapter 10 we will present our new 4D model for joint image reconstruction, spatio-temporal regularization and optimal transport, which reads as follows

Model 7.4.1. *General 4D Reconstruction & Optimal Transport*

$$\begin{aligned} \min_{\rho,\boldsymbol{v}} \; & \int_0^T H_{f(\cdot,t)}\left(K\rho(\cdot,t)\right)\,dt \;+\; \alpha\int_0^T J(\rho(\cdot,t))\,dt \;+\; \frac{\beta}{2}\int_0^T\int_\Omega \rho\,|\boldsymbol{v}|^2\,dx\,dt \qquad (7.1)\\ s.t. \quad & \partial_t\,\rho \;+\; \nabla_\Omega\cdot(\rho\,\boldsymbol{v}) \;=\; 0 \qquad \text{in } Q := \Omega\times[0,T]\\ & \rho \geq 0\ . \end{aligned}$$

On the one hand, this model computes a sequence of reconstructed images ρ from the sequence of given data f in the sense of inverse problems while regarding an optimal transport scheme. On the other hand, the model simultaneously estimates the motion field $\mathbf{v}$ in the sequence, which can be used later on e.g. for registration or tracking purposes. The first term of the objective functional is a data fidelity term H_f in space and time, which depends on the given image sequence f, on a compact operator K and on the desired density sequence ρ. The compact operator K is assumed to have the same structure in all time steps t. Which makes sense for various applications since detector systems like microscopes or tomographs will not change their properties due to time. In analogy to static reconstruction, different data fidelities will result from different noise models using Bayes theory and MAP estimation. The second term of the objective functional is a time dependent regularization term. J can for example be a quadratic regularization term or the total variation. The third term and the constraint is related to the optimal transport scheme with mass conservation we introduced above. The main difference to optimal transport is the fact that we do not have a given start or end density. Initial values for the transport equation automatically arise from the given data of the inverse problem. In addition, we need to add a positivity constraint to the model since we consider density images.

For this general formulation we will study different variants including TV regularization in space and time. We will provide a detailed analysis including existence and

uniqueness proofs. To overcome the large amount of data we will propose two types of numerical realizations based on preconditioning and splitting techniques to facilitate parallelization and efficiency. On the one hand we will present a preconditioned Newton-SQP method with integrated line-search. On the other hand, for the case of TV, we will present a special splitting technique based on inexact Uzawa techniques that is highly parallelized and where each of the resulting substeps in the algorithm is an efficient shrinkage or an efficient DCT inversion, which can additionally be parallelized on GPUs.

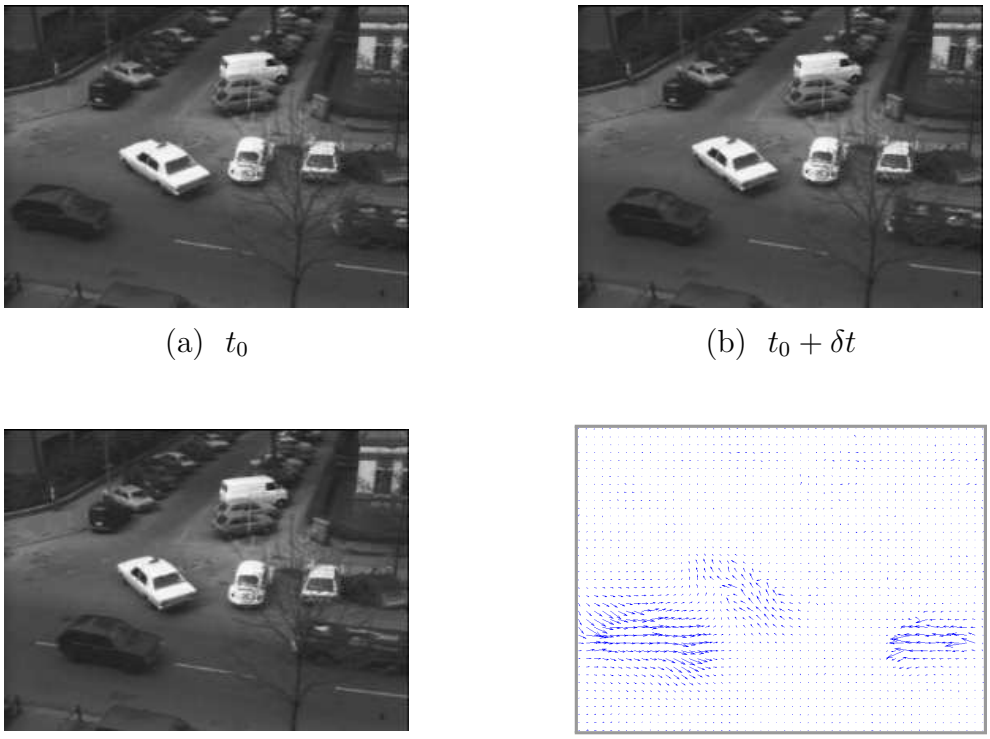
8 Optical Flow and Tracking

In this chapter we will introduce motion fields and general optical flow models for motion estimation in image processing, respectively in imaging. We will introduce the models in a variational setting with data fidelities and regularization techniques. After presenting a global variational framework for optical flow estimation we will concentrate on vectorial TV regularization for 3D motion estimation. We will present a 3D optical flow TV algorithm based on fast split Bregman techniques and show computed tomography (CT) results for heart structure analysis. Furthermore, we will compare optical flow techniques with mass conservation in the sense of compressible and incompressible flows.

8.1 Introduction

Optical flow estimation deals with the determination of visual motion information in image sequences. Let us assume we have a video, in particular a sequence of gray value images,

$$\rho(x,t), \qquad \text{Ort} \;\; x \in \Omega \subset \mathbb{R}^d \,, \quad \text{time} \;\; t \in [0,T] \subset \mathbb{R} \tag{8.1}$$



(a) t_0 (b) $t_0 + \delta t$

(c) $t_0 + 2\ \delta t$ (d) optical flow

Figure 8.1: Hamburg cab sequence and the computed optical flow

Then we search for the optical flow, the velocity field

$$\mathbf{v} := (v_1, v_2, v_3) \quad \text{with } v_1, v_2, v_3 : \Omega \to \mathbb{R}, \tag{8.2}$$

each time between two images $\rho(x, t)$ and $\rho(x, t + \delta t)$. The desired optical flow is a vector field with directional information and velocity and ideally should match the real motion in a scene. Figure 8.1 illustrates a part of the Hamburg cab sequence. An estimation of the optical flow between the two images regarding time t_0 and $t_0 + \delta t$ is depicted in Figure 8.1(d).

In the following we will present a simple and illustrative derivation of variational methods for optical flow as an image processing task. We establish a connection between motion fields and gray value changes, and point out fundamental properties and difficulties with the concept optical flow.

For $t \in [0, T]$ let

$$\rho\ (\cdot, t) : \Omega \to [0, 1] \tag{8.3}$$

be gray value images, more precisely, bounded and measurable functions. For example, imagine you took a sequence of images of a real $3D$ scene. Such as the airplane in Figure 8.1 objects are able to move over time. Let us take a fixed position on the

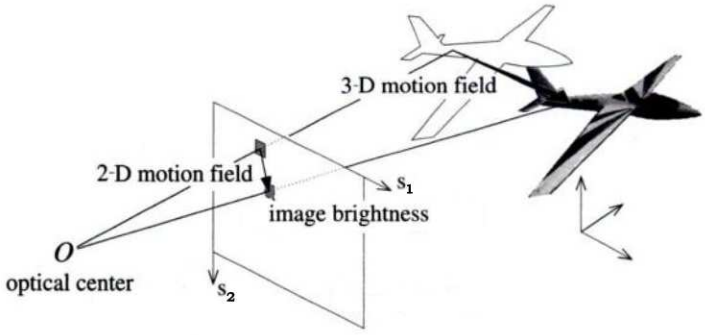


Figure 8.2: Pinhole camera model, from [7], Section 5.1

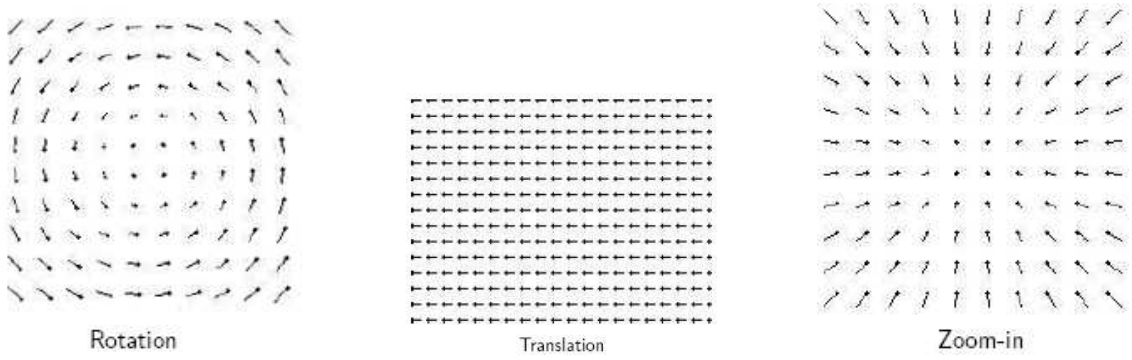
$3D$-surface of the airplane. Following this point over time, we track a $3D$ pathline $S(t)$ with a specific $3D$ position at every time step $t \in [0, T]$. The well-known pinhole camera model explains the projection of a $3D$ scene to the $2D$ image domain. Still following the point on the airplane, the latter now tracks a $2D$ path after projection

$$\mathbf{v}(t) \;=\; (v_1(t), v_2(t)). \tag{8.4}$$

An additional difficulty arises if the person recording the movie itself is sitting in the moving object, for example in a car. In other words, ego motion is an issue in real camera scenes. For instance, if you think of robot or vehicle control, this is an aspect one needs to consider. Starting from a $2D$ path, we obtain the current displacement field, respectively the vectorial velocity of a point on the $2D$ path at time $\hat{t} \in [0, T]$ by the time derivation of the trajectory

$$v_m(\hat{x}) \;:=\; \frac{\partial x}{\partial t}(\hat{t}) \qquad \text{, for } \hat{x} = x(\hat{t}). \tag{8.5}$$

A vector field of $2D$ velocities of visual surface points is called motion field. Very simple examples of motion fields are rotation, translation or zoom. In this context, it is important to lay emphasis on the connection between displacement field, gray value changes and optical flow. The barber's pole illusion is a useful tool to explain this issue. The barber's pole illusion in Figure 8.1 is an counterclockwise rotating cylinder with stripes revolving around similar to a bandage. In this case the (real) displacement field is a translation of visual surface points to the right side. But since we have a visual illusion here, observed gray value variations during the rotation imply an optical flow tending upwards. Gray value changes cannot only be induced by moving objects, but also by camera motion and outside influences like illumination or contrast changes. For instance in registration of MRT images changing contrast can be an issue. We conclude that the optical flow and the displacement field are not identical in general. In the next subsection we will see how different variational



(a) (b) (c)

Figure 8.3: Basic motion fields

models can take into account different motion types and illuminations. The concept of optical flow is an estimation based on a-priori assumptions on intensity changes due to time. In general, the goal is to compute an approximation of the displacement field by time-varying image intensities. A fundamental problem in optical flow is the so called aperture problem. For estimating the optical flow we use spatial and temporal gray values, which are determined via a local operator. For a sufficiently small stencil of the local operator we can only reach a small detail of the object similar to an aperture. Hence we cannot collect enough information on the motion. Simply the components of the motion vector perpendicular to a contour line of the image (i.e. a line with the same intensity value) can be determined. The components parallel to the contour line remain undetected. In Figure 8.1 a view on the aperture gives the impression that the grid moves in bottom-right direction. But in general it could be possible that the grid only moves downwards or only rightwards. Hence the aperture problem is a special case of a correspondence problem. The ambiguity of motion vectors can only be resolved if one can observe "corners" of objects through the aperture, that is the local operator. In the next subsection we will come back to this

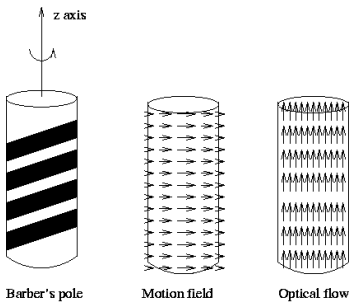


Figure 8.4: Barber's pole illusion

fundamental issue and will verify it mathematically in the sense of ill-posedness and regularity. In this subsection we give an overview of variational methods for optical flow estimation. Starting from the concept of optical flow, we will provide a simple derivation of variational methods and their relation to applications. Based on this, we will deal with the structure of variational methods, in particular data fidelities and regularization terms, similarly to the reconstruction modeling in chapter 5. Major focus will be put on an abstract presentation to give a general overview of existing models for optical flow estimation.

8.2 Model Derivation

A common assumption for estimating optical flow is the preservation of intensity values following motion from one image to the next (brightness constancy). Under the assumption that δt is sufficiently small and that we follow a displacement field $\mathbf{v}_m$, we obtain:

$$\rho\,(x,t) \;=\; \rho\,(x+\mathbf{v}_m(x),\; t+\delta t) \tag{8.6}$$

where $x=(x_1,x_2)^T \in \Omega \subset \mathbb{R}^2$ and where $\mathbf{v}_m : \Omega \to \mathbb{R}^2$ denotes a 2D-velocity field.

A disadvantage of this brightness constancy assumption is the missing invariance under changing illumination. However, in order to guarantee an optimal environment for these assumptions one can create artificial scenes with the following properties:

- We do not have reflections.
- The light-source is far away and punctiform.
- We do not have rotations of objects.

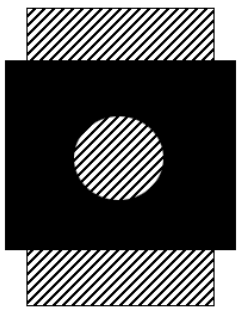


Figure 8.5: Aperture problem

- We only have one source of light.

Obviously these assumptions for brightness constancy are not in line with reality. However, amazingly one can achieve excellent results with this model for several applications. In the course of this subsection we will study more complex assumptions. The latter will show more invariance towards environmental properties and will support various types of motion. To guarantee a good optical flow approximation to the real displacement field it is reasonable in general to assume moderate changes in illumination and distance of camera, light sources or objects. Starting with the brightness assumption, we now want to derive an estimation $\mathbf{v}$ of the $2D$ displacement field $\mathbf{v}_m$. The aim is to find a PDE system for computing the optical flow under this assumption. Let ρ be sufficiently regular, such that we can apply the multidimensional Taylor expansion to the image $\rho(x + \mathbf{v}_m, t + \delta t)$, whose grid has been transformed by the displacement field $\mathbf{v}_m$.

$$\begin{aligned} \rho(x,t) &\overset{(8.6)}{=} \rho(x + \mathbf{v}_m, t + \delta t) \\ &\overset{\text{Taylor}}{=} \rho(x,t) + {\mathbf{v}_m}^T \nabla_2 \rho(x,t) + \rho_t(x,t) \quad + \quad O(D^2\rho) \end{aligned} \tag{8.7}$$

with $\nabla_2 \rho = (\rho_{x_1}, \rho_{x_2})^T$ as spatial derivative and ρ_t as temporal derivative. Thus, considering the brightness constancy approximatively implies a basic constraint for computing the optical flow $\mathbf{v}$:

$$\mathbf{v}(x,t)^T \nabla_2 \rho(x,t) + \rho_t(x,t) = 0. \tag{8.8}$$

We obtain a gradient condition for the optical flow, the so-called *optical flow constraint* (OFC). The underlying operator

$$\frac{D}{Dt} := \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$$

is called Lagrange operator oder material derivative and is often used in fluid mechanics and classical mechanics. This point of view is related to compressible and incompressible flows, see Section 8.7.

The optical flow constraint relates the optical flow $\mathbf{v}$ to spatial and temporal derivatives of the density image ρ. Since $\mathbf{v}$ is a vector field

$$\mathbf{v} := (v_1, v_2)^T \ : \ \mathbb{R}^2 \supset \Omega \to \mathbb{R}^2 \ ,$$

we seek a solution of **one** functional equation with **two** unknowns. We have an under-determined system of equations and $\mathbf{v}$ cannot be determined uniquely from the optical flow constraint itself. The gradient constraint equation restricts the optical flow to a straight line. The underdetermination of the equation is related to the

aperture problem we introduced in the previous subsection. Under this assumptions we speak of an ill-posed problem. Following HADAMARD a problem is well-posed if the following conditions are fulfilled:

- There exists a solution of the problem. (existence)
- The solution is unique. (uniqueness)
- The solution depends continuously on the given data. (stability)

Obviously, the uniqueness condition is broken in our case. As illustrated in Figure

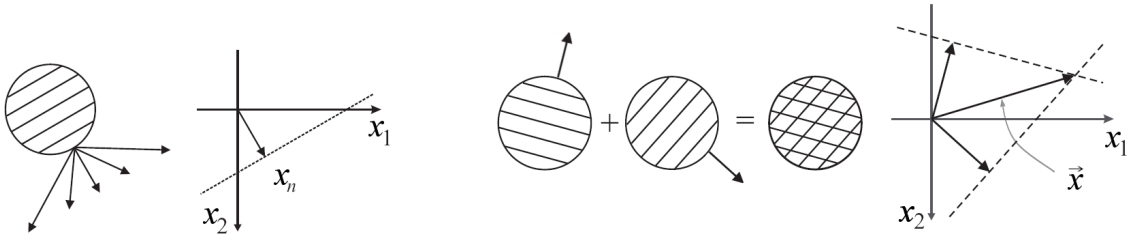


Figure 8.6: Missing uniqueness due to the aperture problem

8.6, we can only obtain the components of $\mathbf{v}$ in direction of $\nabla\rho$, that is the normal to the contour lines of the image,

$$\mathbf{v}_n = \frac{-\rho_t}{|\nabla_2\rho|}\frac{\nabla_2\rho}{|\nabla_2\rho|} \ . \tag{8.9}$$

To penalize deviations from the optical flow constraint in 8.8, variational methods are based on minimizing a *data fidelity* with the L^p-norm of this equation,

$$(\mathbf{v}^T \ \nabla_2\rho + \rho_t)^2 \ . \tag{8.10}$$

The locality of the neighborhood does not yield a unique solution for the optical flow. Hence we seek a vector field $\mathbf{v}$ which minimizes the error in the data fidelity. If one allows more than two or more gradient directions, then the various intersecting "condition lines" can form a unique condition for the flux vector $\mathbf{v}$. This leads to higher order data fidelities, which we will formalize and classify in a global framework. Similarly to ill-posed inverse problems, here one can overcome the locality problem, i.e. the aperture problem, by introducing regularization terms. The latter propagate information in a sufficiently global sense (sufficiently smooth solutions). In Subsection 8.4 we will study flow regularization terms in further detail. The following part deals with the structure of functionals for variational problems in optical flow estimation. To simplify notation we introduce multiindices:

Definition 8.2.1 (Multiindex). *Let $\rho : \mathbb{R}^d \supset \Omega \to \mathbb{R}$. For $k \in \mathbb{N}$ an element of the set*

$$A_k \;:=\; \{a = (a_1, ..., a_d),\; a_i \in \mathbb{N},\; |a| = \sum_{i=1}^{d} a_i = k\}$$

is called multiindex of order k. For $a \in A_k$ we write

$$D^a\rho \;:=\; D_1^{a_1}...D_d^{a_d}\rho \;=\; \frac{\partial^{|a|}}{\partial {x_1}^{a_1}...\partial {x_d}^{a_d}}\;.$$

Furthermore, let $\nabla^k\rho \;:=\; (D^a\rho)_{a\in A_k}$, i.e. $\nabla^k\rho$ contains all partial derivatives of order k of the function ρ.

To compute the optical flow $\mathbf{v} = (v_1, v_2, v_3)^T$ we minimize the following objective functional for a given density ρ:

$$\min_{\mathbf{v}} \underbrace{\int_\Omega m(\nabla^k\rho,\; \mathbf{v})\; dx}_{\text{data fidelity}} \;+\; \mu \underbrace{\int_\Omega r(\nabla_2\rho,\; \nabla_2 v_1,\; \nabla_2 v_2)\; dx}_{\text{regularization term}} \tag{8.11}$$

$x := (x_1, x_2)^T \in \Omega$ space

$\nabla_2 := \nabla_{(x_1,x_2)} = (\partial_{x_1}, \partial_{x_2})^T$.

In general, a data fidelity term $m(\cdot,\cdot)$ consists of (mixed) partial derivatives of gray value intensities with order k, which are related to $\mathbf{v}$ via a constraint. In other words, the minimization of a data fidelity penalizes deviations from a specific constancy assumption. We use this general formulation with order k to include various types or data models in the next subsection. However, a regularization term $r(\cdot,\cdot)$ penalizes deviations from the smoothness of the flux. In subsection 8.4 and 8.5 we will present various flow regularization techniques including total variation. Similarly to reconstruction problems the parameter α is a weighting parameter to control the influence of the two components. Here we have $d = 2$, i.e. we have multiindices of the form $a = (a_1, a_2) \in \mathbb{N}_0^2$ and the following holds:

$$D^a\rho \;:=\; D_1^{a_1}D_2^{a_2}\rho \;=\; \frac{\partial^{|a|}\rho}{\partial {x_1}^{a_1}\;\partial {x_2}^{a_2}}. \tag{8.12}$$

Derivations of an order smaller than 2 we can reach by corresponding zero values in the multiindex. For example one can define the first partial derivatives via unit vectors as multiindices:

$$D_i\rho \;=\; D^{e_i}\rho \;=\; \frac{\partial\rho}{\partial s_i}\;. \tag{8.13}$$

With respect to $4D$ image reconstruction in Chapter 10, it is useful to consider the optical flow problem as a time dependent problem. This extension to $3D + 1D$ motion estimation means the following: Instead of considering only two D-dimensional

densities and their derivatives, one considers a whole sequence of D-dimensional densities and additionally temporal derivatives of the flux. This extension to space-time motion estimation can simply be incorporated into our framework above. Instead of a spatial domain $\Omega \subset \mathbb{R}^d$ we pass over to a space-time domain $Q := \Omega \times [t_0, t_1]$. For a given sequence of density images $\rho : \Omega \times [t_0, t_1] \to \mathbb{R}$ variational problems for the optical flow in general have the following structure:

$$\min_{\mathbf{v}} \underbrace{\int_Q m(\nabla^k \rho,\ \mathbf{v})\ d\tilde{x}}_{\text{data fidelity}} \ +\ \alpha \underbrace{\int_Q r(\nabla_3 \rho,\ \nabla_3 v_1,\ \nabla_3 v_2)\ d\tilde{x}}_{\text{regularization term}} \tag{8.14}$$

$$\tilde{x} := (x_1, x_2, t)^T \in \Omega \times [t_0, t_1] =: Q \quad \text{space-time}$$
$$\nabla_3 := \nabla_{(x_1,x_2,t)} = (\partial_{x_1}, \partial_{x_2}, \partial_t)^T.$$

Here we have $d = 3$, i.e. the multiindices are extended by one dimension to $a = (a_1, a_2, a_3) \in \mathbb{N}_0^3$, and we obtain:

$$D^a \rho \ :=\ D_1^{a_1} D_2^{a_2} D_3^{a_3} \rho \ =\ \frac{\partial^{|a|} \rho}{\partial {x_1}^{a_1}\ \partial {x_2}^{a_2}\ \partial t^{a_3}}\ . \tag{8.15}$$

In the literature there are some papers on optical flow estimation that present improved quality achieved by space-time modeling. However, in such a case, one has to get along with an additional dimension and more data computationally. In the last chapter on $4D$ image reconstruction we use a space-time motion model in terms of optimal transport. For simplicity reasons, we will restrict ourselves to the motion computation between two D-dimensional density images for the rest of this chapter, i.e. we write

$$\rho(x) \text{ for } \rho(x, 0) \quad \text{and} \quad \nabla v_i \text{ for } \nabla_2 v_i,\ i = 1, 2\ .$$

As a simple example for optical flow estimation, we consider the variational method by Horn and Schunck, see [103, pages 81-87],

$$\min_{\mathbf{v}} \int_\Omega (\ \rho_{v_1}(x)\ v_1(x) + \rho_{v_2}(x)\ v_2(x) + \rho_t(x)\ )^2\ dx \\ + \int_\Omega |\nabla v_1(s)|^2 + |\nabla v_2(x)|^2\ dx\ . \tag{8.16}$$

This is a global method with brightness constancy as data fidelity m and a simple quadratic, homogeneous regularization term r if one uses the general notation from above.

8.3 Data Fidelities

In this section we concentrate on the structure of data fidelity term for motion estimation.

$$\min_{\mathbf{v}} \int_{\Omega} m(\nabla^k \rho,\ \mathbf{v})\ dx\ +\ \text{"Regularizer"} \tag{8.17}$$

Analogous to the derivation of the optical flow constraint $\tilde{\mathbf{v}}^T \nabla_3 \rho = 0$, starting from brightness constancy, we can design gradient constraints of higher order. For this purpose, one assumes the preservation of intensity derivatives from one frame to another while following the motion. For example the preservation of intensity derivatives of first order (gradient constancy),

$$\begin{aligned} \rho_{x_1}\ (x,t) &\stackrel{!}{=} \rho_{x_1}\ (x + \mathbf{v}_m(s),\ t + \delta t)\ , \\ \rho_{x_2}\ (x,t) &\stackrel{!}{=} \rho_{x_2}\ (x + \mathbf{v}_m(s),\ t + \delta t)\ , \end{aligned} \tag{8.18}$$

leads analogously to the following PDEs

$$\begin{aligned} \mathbf{v}(x,t)^T\ \nabla \rho_{x_1}(x,t) + \rho_{x_1,t}(x,t) = 0\ , \\ \mathbf{v}(x,t)^T\ \nabla \rho_{x_2}(x,t) + \rho_{x_2,t}(x,t) = 0\ . \end{aligned} \tag{8.19}$$

Hence, the corresponding data fidelities are given by the L^2 norm of the left hand side of the stationary equation in (8.19). Regarding the ambiguity in the aperture problem (8.6) this is an advantage as we have more constraint equations available for the estimation. In the following table we list different data fidelities and their corresponding constancy assumptions. In the following let $\tilde{\mathbf{v}} = (v_1, v_2, 1)^T$. Actually the

Type of Constancy	Data Fidelity	Type of Motion
gradient	$\sum_{i=1}^{2} (\tilde{\mathbf{v}}^T \nabla_3 \rho_{x_i})^2$	translation, zoom-out, slow rotation
2nd derivatives	$\sum_{i=1}^{2} \sum_{j=1}^{2} (\tilde{\mathbf{v}}^T \nabla_3 \rho_{x_i x_j})^2$	translation, zoom-out, slow rotation

Table 8.1: Motion depending constancy assumptions and their data fidelities

terms in Table 8.2 are invariant under all motion types. Theoretically, data fidelities of higher order can be used to take into account every differentiable image property. However, using derivatives of higher order can cause more sensitivity difficulties due to disturbances. For example, the gradient constancy implies that the motion field should not have deformations of first order. Conversely, the brightness constancy assumption only requires a certain smoothness due to the transition between two

Type of Constancy	Data Fidelity	Type of Motion
absolute value of gradient	$(\tilde{\mathbf{v}}^T \nabla_3 \lvert\nabla_2 \rho\rvert)^2$	all
Laplacian	$(\tilde{\mathbf{v}}^T \nabla_3 \, \Delta\rho)^2$	all

Table 8.2: Motion invariant constancy assumptions and their data fidelities

frames. Considering different data fidelities reveals a common structure. Thus, similarly to combining different noise models in one image reconstruction model, we can combine different constancy assumptions and their data fidelities. This leads to a notation via motion tensors, as they have been proposed in [28] the first time. For constancy assumptions $p_1, .., p_n$ with weights $\gamma_1, .., \gamma_n \in \mathbb{R}$ we obtain

$$\sum_{i=1}^{n} \gamma_i (\tilde{\mathbf{v}}^T \nabla_3 p_i)^2 \;=\; \tilde{\mathbf{v}}^T \left(\sum_{i=1}^{n} \gamma_i \nabla_3 p_i (\nabla_3 p_i)^T \right) \tilde{\mathbf{v}} \;=\; \tilde{\mathbf{v}}^T \underbrace{M(\nabla_3 p_1, .., \nabla_3 p_n)}_{\text{motion tensor}} \tilde{\mathbf{v}} \; . \tag{8.20}$$

A weighted combination of data fidelities is interesting because they can be written in a unified **quadratic form**. This notation offers a whole framework of different possible data fidelities.

Type of Constancy p_i	Motion Tensor $M(\nabla_3 p_1, .., \nabla_3 p_n)$
brightness	$M(\nabla_3 \rho) \;=\; \nabla_3 \rho (\nabla_3 \rho)^T$
gradient	$M(\nabla_3 \rho_{x_1}, \nabla_3 \rho_{x_2}) \;=\; \sum_{i=1}^{2} (\nabla_3 \rho_{x_i})(\nabla_3 \rho_{x_i})^T$
2nd derivatives	$M(\nabla_3 \rho_{x_1 x_1}, .., \nabla_3 \rho_{x_2 x_2}) \;=\; \sum_{i=1}^{2} \sum_{j=1}^{2} (\nabla_3 \rho_{x_i x_j})(\nabla_3 \rho_{x_i x_j})^T$
absolute value of gradient	$M(\nabla_3 \lvert\nabla_2 \rho\rvert) \;=\; \frac{(\rho_{x_1} \nabla_3 \rho_{x_1} + \rho_{x_2} \nabla_3 \rho_{x_2})(\rho_{x_1} \nabla_3 \rho_{x_1} + \rho_{x_2} \nabla_3 \rho_{x_2})^T}{\rho_{x_1}^2 + \rho_{x_2}^2}$
Laplacian	$M(\nabla_3 \Delta\rho) \;=\; (\nabla_3 \sum_{i=1}^{2} \rho_{x_i x_i})(\nabla_3 \sum_{i=1}^{2} \rho_{x_i x_i})^T$

Table 8.3: Constancy assumptions and their corresponding motion tensors

To classify the previously introduced constancy assumptions in this notation, we list the motion tensors regarding specific data fidelities in Table 8.3.

Example:
The motion tensor for preserving the absolute value of the gradient of intensities can be computed in the following way:

$$\nabla_3|\nabla_2\rho| \;=\; \frac{\nabla_3(\rho_{x_1}^2+\rho_{x_2}^2)}{2|\nabla_2\rho|} \overset{\text{linear}}{=} \frac{\nabla_3\rho_{x_1}^2+\nabla_3\rho_{x_2}^2}{2|\nabla_2\rho|} \overset{\text{chain rule}}{=} \frac{\rho_{x_1}\nabla_3\rho_{x_1}+\rho_{x_2}\nabla_3\rho_{x_2}}{|\nabla_2\rho|}\;,$$

which implies the quadratic form

$$\tilde{\mathbf{v}}^T\left((\nabla_3|\nabla_2\rho|)(\nabla_3|\nabla_2\rho|)^T\right)\tilde{\mathbf{v}} \;= \underbrace{\tilde{\mathbf{v}}^T\left(\frac{(\rho_{x_1}\nabla_3\rho_{x_1}+\rho_{x_2}\nabla_3\rho_{x_2})(\rho_{x_1}\nabla_3\rho_{x_1}+\rho_{x_2}\nabla_3\rho_{x_2})^T}{\rho_{x_1}^2+\rho_{x_2}^2}\right)\tilde{\mathbf{v}}}_{=M(\nabla_3|\nabla_2\rho|)}\;.$$

8.3.1 Robustness

Before going over to regularization terms, we will concentrate on robustness of data fidelities. We suggest two principles for reducing the influence of outliers in the data. These techniques are known as low-pass filters. High frequencies are filtered, whereas low frequencies in the data are allowed to pass.
The first approach we consider is the well-known method for optical flow computation by Lucas and Kanade [121], [120]. In this case the quadratic form

$$\min_{\mathbf{v}}\; \tilde{\mathbf{v}}^T M_\sigma(\nabla_3\rho)\;\tilde{\mathbf{v}} \tag{8.21}$$

is minimized using a filter $M_\sigma(\nabla_3\rho) := K_\sigma * (\nabla_3\rho\;(\nabla_3\rho)^T)$ with a Gaussian kernel K_σ (scale $\sigma > 0$). In comparison to global variational methods we do not minimize an integral over the whole space but only locally. A motion tensor is filtered locally by a Gaussian kernel K_σ. A computational advantage of simple Gaussian filtering results from the fact that it can be realized via a simple multiplication in a Fourier space (convolution theorem). A local method has the advantage of being robust towards outliers in the data, but has the drawback that resulting optical flow fields are not dense. However global methods compute dense vector fields but are more sensitive towards noise. Bruhn, Weickert and Schnörr suggested a combination of local and global methods [29] attaining robust variational methods.

$$\min(\mathbf{v})\;\underbrace{\int_\Omega \tilde{\mathbf{v}}^T M_\sigma(\nabla_3\rho)\;\tilde{\mathbf{v}}\;dx}_{\text{robust data fidelity}}\;+\text{"Regularizer"} \tag{8.22}$$

In addition, a computational advantage results from the fact, that the convolution is only related to single partial density derivatives. Hence, the convolution can be done in a preprocessing step. Now we proceed to the second component of variational methods for optical flow estimation, the regularization techniques.

8.4 Regularization

In this section deals with the vector field regularization in variational methods for optical flow computation. Consider a variational problem of the form

$$\min_{\mathbf{v}} \text{ "Data fidelity"} \ + \ \alpha \int_{\Omega} r(\nabla_2 \rho, \ \nabla v_1, \ \nabla v_2) \ dx \ .$$

As described in 8.3, with data fidelities, e.g. the term $(\mathbf{v}^T \ \nabla_2 \rho + \rho_t)^2$ (brightness constancy, OFC), the optical flow cannot be determined uniquely in general. We have to make use of additional prior information to work against the aperture problem and to ensure well-posedness. Hence, as common for variational methods in imaging, the optical flow methods introduced here are based on regularization. In general one assumes a certain smoothness of the optical flow. The parameter α is a control parameter weighting the influence of data fidelity, respectively regularization term. r denotes a regularization term and usually depends on derivatives of the optical flow field $\mathbf{v}$ or possibly, in addition, from intensity derivatives. In the following we study different types of regularization terms. The methods of Horn and Schunck (8.16) is a prototype for homogeneous regularization. Their smoothness constraints are based on quadratic regularization terms, i.e.

$$\begin{aligned} &\text{spatial homogeneous smoothness :} && r_H \ := \ |\nabla v_1|^2 + |\nabla v_2|^2 \\ &\text{spatio-temporal homogeneous smoothness :} && r_{H3D} \ := \ |\nabla_3 v_1|^2 + |\nabla_3 v_2|^2 \ . \end{aligned}$$

A lot of different regularization techniques are related to or can be motivated by diffusion processes. A diffusion process is part of a diffusion-reaction system and describes the temporal evolution process, which is in our case reflects the process of smoothing the unknown velocity field. One can think of a scale space method. The Euler-Lagrange equations for such a problem can be seen as the stationary elliptic equation of the corresponding parabolic equation. Now we will have a closer look on the diffusion part. For the homogeneous method above, the corresponding diffusion process reads as follows:

$$\frac{\partial v_i}{\partial \tau} \ = \ \nabla \cdot \ (Id \ \nabla v_i), \qquad i = 1, 2 \tag{8.23}$$

where τ denotes a time step (diffusion time) and Id the identity matrix. The smoothness is called homogeneous since the diffusivity has value 1 everywhere.

8.4.1 Image-Driven Regularization

Tracking and segmentation of moving objects is an interesting aspect in motion computation. For example one can think of pedestrian motion, traffic motion or

cell migration. In such applications one is usually more interested in "motion edges" (motion segmentation) rather than "density edges" (density segmentation). In some way it is reasonable to see the discontinuities in the motion field as a subset of the discontinuities in the density. An easy way to reduce smoothing at motion edges (discontinuities in the optical flow) is the introduction of a weighting function g in the homogeneous regularization term

$$r_{B\rho}(\nabla_2\rho,\ \nabla v_1,\ \nabla v_2)\ :=\ g(|\nabla_2\rho|^2)\ \sum_{i=1}^{2}|\nabla v_i|^2\ . \tag{8.24}$$

Let the weighting function g be a positive and monotone decreasing function which takes small values at intensity edges (discontinuities in density images). This type of regularization has been suggested by Alvarez et al. [2]. The underlying diffusion process reads as follows

$$\frac{\partial v_i}{\partial \tau}\ =\ \nabla\cdot\ (g(|\nabla_2\rho|^2)\ \nabla v_i), \qquad i=1,2\ . \tag{8.25}$$

Obviously, the diffusion is dependent on intensity edges, however it implies a direction independent smoothing (isotropic diffusion). Thus, the result in 8.24 is generated by an inhomogeneous, image-driven and isotropic regularization term. Charbonnier et al. [44] proposed the function g as :

$$g(s^2) := \frac{1}{\sqrt{1+\frac{s^2}{\delta_s^2}}}\ , \tag{8.26}$$

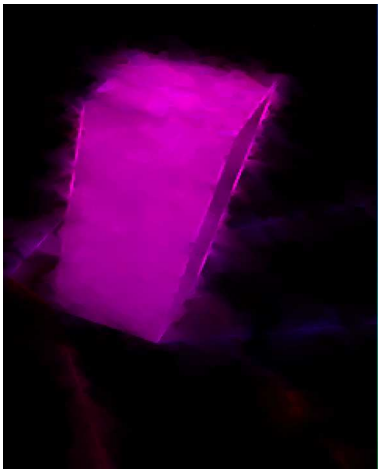
where δ_s is a relaxation parameter. An image-driven isotropic regularization term avoids smoothing near intensity edges, however it does not allow for the direction of intensity derivatives, see Figure 8.4.1 (a). The basic idea of an anisotropic method is to avoid smoothing across edges, whereas smoothing along edges is preferred. In motion estimation one of the first techniques of this type has been suggested in [130]. They defined regularization terms in the following way

$$r_{BA}(\nabla_2\rho,\ \nabla v_1,\ \nabla v_2)\ :=\ \sum_{i=1}^{2}(\nabla v_i)^T\ Dif_{NE}(\nabla_2\rho)\ \nabla v_i\ , \tag{8.27}$$

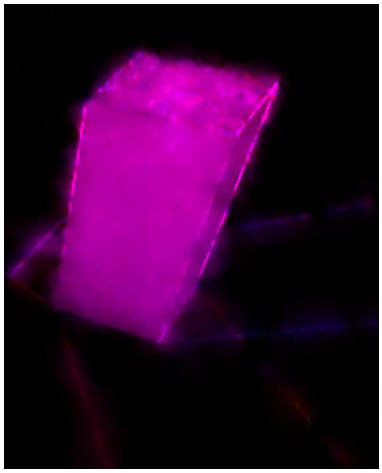
$$\text{where}\qquad Dif_{NE}(\nabla_2\rho)\ :=\ \frac{1}{|\nabla\rho|^2+2\epsilon_s^2}\begin{pmatrix}\rho_{x_2}^2+\epsilon_s^2 & -\rho_{x_1x_2}\\ -\rho_{x_1x_2} & \rho_{x_1}^2+\epsilon_s^2\end{pmatrix}$$

$\epsilon_s > 0$ is a parameter, which ensures the regularity of the matrix.

Figure 8.7: image-driven regularization in the *new-marbled-block*-sequence



(a) image-driven, isotropic



(b) image-driven, anisotropic

The diffusion process for this kind of regularization is given by:

$$\frac{\partial v_i}{\partial \tau} \;=\; \nabla \cdot \; (Dif_{NE}(\nabla_2 \rho)\; \nabla v_i), \qquad i = 1, 2\ . \tag{8.28}$$

Taking a look at the eigenvectors and the corresponding eigenvalues of Dif_{NE} one can read out the desired behavior at edges considering $|\nabla_2 \rho| \to 0$ and $|\nabla_2 \rho| \to \infty$. The described properties are presented in Figure 8.4.1 (b). Similar properties are known from anisotropic diffusion filters [186]. Homogeneous and image-driven isotropic regularizations are special cases of (8.27), simply choose

$$Dif_{NE}(\nabla_2 \rho) \;:=\; \rho \qquad \text{resp.} \qquad Dif_{NE}(\nabla_2 Id) \;:=\; g(|\nabla_2 \rho|^2) Id\ ,$$

where Id denotes the identity matrix.

8.4.2 Flow-Driven Regularization

In image-driven regularization the smoothness of motion fields is influenced by intensity gradients. Particularly for objects with numerous and intense textures, e.g. see the surface of the block in 8.4.1, this means we have to expect a textured motion estimation. We obtain much more intensity edges than motion edges. If one is primarily interested in discontinuities of the optical flow, for example in tracking applications, than this effect is not desired. To overcome this "drawback" one can go over to flow-driven regularization terms. Flow-driven regularization methods predominantly take into account discontinuities in the motion field. Analog to

the previous subsection we distinguish between isotropic and anisotropic methods. Isotropic flow-driven methods are roughly based on the idea of adding a filter function to the homogeneous regularization term, eliminating outliers in $|\nabla x_1|^2 + |\nabla x_2|^2$,

$$r_{F\rho}(\nabla v_1, \nabla v_2) \;:=\; \Psi(|\nabla v_1|^2 + |\nabla v_2|^2) \;, \tag{8.29}$$

where Ψ denotes a positive and increasing function with values in $\mathbb{R}$.
A prototype for Ψ is

$$\Psi(s^2) = \sqrt{s^2 + \epsilon} \;, \tag{8.30}$$

which leads to an approximated isotropic total variation regularization of the flow with relaxation parameter $\epsilon > 0$. Analog to scalar total variation this model is related to lagged diffusivity regarding the additional parameter ϵ. In general one can expect a nearly piecewise constant behavior in each component of the velocity field. Analog to robust data fidelity terms, we introduced in section (8.3.1), penalizing with such a function Ψ can be seen as a statistically robust error measure [104]. The corresponding diffusion process reads as follows

$$\frac{\partial v_i}{\partial \tau} \;=\; \nabla \cdot \; (\Psi'(|\nabla v_1|^2 + |\nabla v_2|^2) \; \nabla v_i), \qquad i = 1, 2 \;. \tag{8.31}$$

The diffusivity decreases with its argument. Hence smoothing at discontinuities of the optical flow is avoided. An alternative approximation of TV regularization was proposed in [188]. Rotationally invariant approximations can be found in [46], [56] or [113].

The second class of flow-driven vector field regularization is the anisotropic one. In the isotropic case the non-quadratic real-valued function Ψ penalizes the absolute value of the flow derivatives. Here we extend the function Ψ to a matrix valued transformation Ψ_m. We can apply this transformation on

$$\sum_{i=1}^{2} \nabla v_i \; (\nabla v_i)^T \;, \tag{8.32}$$

which contains additional directional information. The transformation can be defined in the following way

$$\Psi_m(A) \;:=\; \sum_{i=1}^{2} \Psi(\sigma_i) \; \omega_i \omega_i^T \;, \tag{8.33}$$

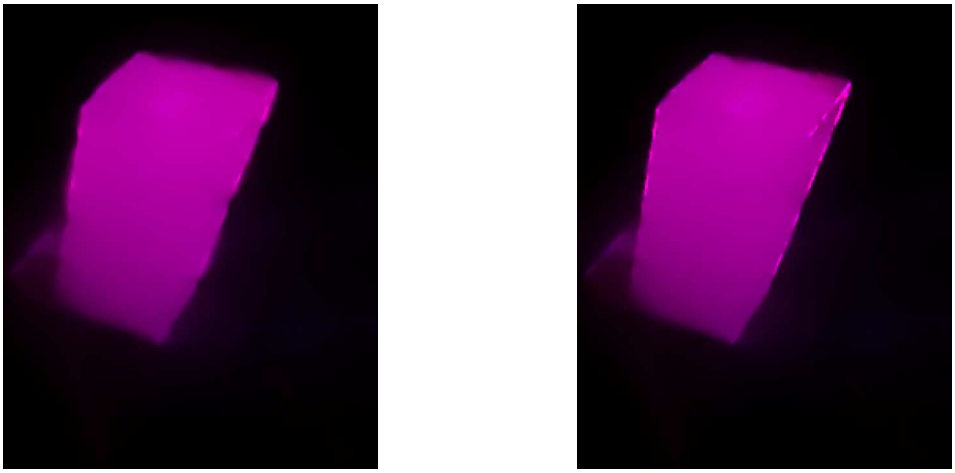
where ω_i are orthogonal eigenvectors and σ_i are corresponding eigenvalues for a symmetric matrix $A \in \mathbb{R}^2 \times \mathbb{R}^2$. The matrix (8.32) is a symmetric positive semidefinite matrix. The application of Ψ_m yields a partition of the eigenvalues. Using the trace

of Ψ_m we obtain the sum of eigenvalues (8.32). With

$$r_{FA}(\nabla v_1, \nabla v_2) \;:=\; trace\,(\Psi_m(\sum_{i=1}^{2} \nabla v_i\ (\nabla v_i)^T)) \tag{8.34}$$

we finally obtain a regularization term which adjusts to the local orientation of the optical flow. Following [187], this type of regularization implies the diffusion process

Figure 8.8: flow-driven regularization in the *new-marbled-block*-sequence



(a) flow-driven, isotropic

(b) flow-driven, anisotropic

$$\frac{\partial v_i}{\partial \tau} \;=\; \nabla \cdot\ (Dif(\nabla v_1, \nabla v_2)\ \nabla v_i), \qquad i = 1, 2, \tag{8.35}$$

$$\text{where}\quad Dif(\nabla v_1, \nabla v_2) \;=\; \Psi'_m(\sum_{i=1}^{2} \nabla v_i\ (\nabla v_i)^T)\ .$$

Analog to the image-driven case, the additional consideration of local orientation leads to a smoothing behavior along motion edges but not across motion edges. For further insight on anisotropic filters in image processing we refer to [186]. Figure 8.8 illustrates results using flow-driven isotropic and anisotropic regularization. In Table 8.4 we give an overview of the four described regularization classes.

Table 8.4: Classes of Flow Regularization Techniques

Regularization	**Smoothing Term**
homogeneous	$\sum_{i=1}^{2} \lvert\nabla v_i\rvert^2$
image-driven, isotropic	$g(\lvert\nabla_2 \rho\rvert^2) \sum_{i=1}^{2} \lvert\nabla v_i\rvert^2$
image-driven, anisotropic	$\sum_{i=1}^{2} (\nabla v_i)^T Dif(\nabla_2 \rho)\ \nabla v_i$
flow-driven, isotropic	$\Psi(\sum_{i=1}^{2} \lvert\nabla v_i\rvert^2)$
flow-driven, anisotropic	$trace\ (\Psi_m(\sum_{i=1}^{2} \nabla v_i\ (\nabla v_i)^T))$

8.5 3D Optical Flow-TV

In this section we consider an optical flow model with the standard brightness constancy assumption and vector field regularization with total variation. We will apply the Split Bregman algorithm to obtain a fast computational method.

8.5.1 Model: Optical Flow-TV

Model 8.5.1.

$$\min_{\boldsymbol{v}=(v_1,v_2,v_3)} \left\{ \frac{1}{2} \int_\Omega (\rho_t + \boldsymbol{v} \cdot \nabla \rho)^2 \, dx \ + \ \alpha \int_\Omega |\nabla v_1| + |\nabla v_2| + |\nabla v_3| \ dx \right\}$$

This variational problem is based on a data fidelity and a vector field regularization term that penalizes deviations from the smoothness of the optical flow. The data fidelity is based on the optical flow constraint regarding the brightness constancy assumption, as introduced in section 8.3 above. The regularization term can be classified as anisotropic and flow-driven due to the terminology of the previous section. In addition, we have another isotropy/anisotropy choice for each of the l^p norms of the vector field components, i.e. analoguous to scalar TV regularization we can choose for instance

$$|\nabla v_i|_{l^2} \qquad \text{or} \qquad |\nabla v_i|_{l^1} \ ,$$

for the isotropic, respectively the anisotropic case.

In the following we want to applied a convex splitting algorithm to find a solution of this variational problem for 3D optical flow TV computation. First of all we have to decouple the functional similarly to the idea of splitting techniques we introduced in Chapter 4. By adding new (artificial) primal variables z_1, z_2, z_3 we obtain a constrained optimization problem that is equivalent to Model 8.5.1:

$$\min_{\mathbf{v}=(v_1,v_2,v_3)} \left\{ \frac{1}{2} \int_\Omega (\rho_t + \mathbf{v} \cdot \nabla \rho)^2 \, dx \; + \; \alpha \int_\Omega |z_1| + |z_2| + |z_3| \; dx \right\}$$
$$\text{s.t.} \quad z_1 = \nabla v_1, \quad z_2 = \nabla v_2, \quad z_3 = \nabla v_3 \; .$$

The augmented Lagrangian for this constrained optimization problem reads as follows

$$\begin{aligned} & L^\mu(v_1, v_2, v_3, z_1, z_2, z_3;\; p_1, p_2, p_3) \\ = & \frac{1}{2} \int_\Omega (\rho_t + \mathbf{v} \cdot \nabla \rho)^2 \; + \; \alpha \left(|z_1| + |z_2| + |z_3| \right) \, dx \; + \; \langle p_1, \nabla v_1 - z_1 \rangle \; + \; \frac{\mu}{2} \left\| \nabla v_1 - z_1 \right\|_2^2 \\ & + \; \langle p_2, \nabla v_2 - z_2 \rangle \; + \; \frac{\mu}{2} \left\| \nabla v_2 - z_2 \right\|_2^2 \; + \; \langle p_3, \nabla v_3 - z_3 \rangle \; + \; \frac{\mu}{2} \left\| \nabla v_3 - z_3 \right\|_2^2 \; , \end{aligned}$$

where μ is a penalty parameter for the quadratic relaxation terms, and where the new dual variables p_1, p_2, p_3 are Lagrange multipliers corresponding to the three constraints.

8.5.2 Algorithm: Split Bregman Optical Flow-TV

After a suitable decoupling of a variational problem, the main idea of Split Bregman (respectively of *alternating direction method of multipliers* (ADMM) and Douglas-Rachford-Splitting (DRS)) is to alternatively minimize the augmented Lagrangian regarding primal and dual variables. In the case of 3D optical flow with total variation we obtain the following primal-dual update scheme:

Primal updates:

$$\mathbf{v}^{k+1} = \underset{\mathbf{v}=(v_1,v_2,v_3)}{\arg\min} \; L^\mu(v_1, v_2, v_3, z_1^k, z_2^k, z_3^k;\; p_1^k, p_2^k, p_3^k) \tag{I}$$

$$z_1^{k+1} = \underset{z_1}{\arg\min} \; L^\mu(v_1^{k+1}, v_2^{k+1}, v_3^{k+1}, z_1, z_2^k, z_3^k;\; p_1^k, p_2^k, p_3^k) \tag{II}$$

$$z_2^{k+1} = \underset{z_2}{\arg\min} \; L^\mu(v_1^{k+1}, v_2^{k+1}, v_3^{k+1}, z_1^{k+1}, z_2, z_3^k;\; p_1^k, p_2^k, p_3^k) \tag{II}$$

$$z_3^{k+1} = \underset{z_3}{\arg\min} \; L^\mu(v_1^{k+1}, v_2^{k+1}, v_3^{k+1}, z_1^{k+1}, z_2^{k+1}, z_3;\; p_1^k, p_2^k, p_3^k) \tag{IV}$$

Dual updates:

$$
\begin{aligned}
p_1^{k+1} &= p_1^k \;+\; \mu\, \partial_{p_1} L^\mu(v_1^{k+1}, v_2^{k+1}, v_3^{k+1}, z_1^{k+1}, z_2^{k+1}, z_3^{k+1};\; p_1, p_2^k, p_3^k) \\
&= p_1^k \;+\; \mu \left(\nabla v_1^{k+1} - z_1^{k+1}\right) && \text{(V)} \\
p_2^{k+1} &= p_2^k \;+\; \mu\, \partial_{p_2} L^\mu(v_1^{k+1}, v_2^{k+1}, v_3^{k+1}, z_1^{k+1}, z_2^{k+1}, z_3^{k+1};\; p_1^{k+1}, p_2, p_3^k) \\
&= p_2^k \;+\; \mu \left(\nabla v_2^{k+1} - z_2^{k+1}\right) && \text{(VI)} \\
p_3^{k+1} &= p_3^k \;+\; \mu\, \partial_{p_3} L^\mu(v_1^{k+1}, v_2^{k+1}, v_3^{k+1}, z_1^{k+1}, z_2^{k+1}, z_3^{k+1};\; p_1^{k+1}, p_2^{k+1}, p_3) \\
&= p_3^k \;+\; \mu \left(\nabla v_3^{k+1} - z_3^{k+1}\right) \;. && \text{(VII)}
\end{aligned}
$$

The dual updates can be computed very efficiently since we only have to apply gradients. We now take a closer look at the primal iterates. Without affecting minimizers, we obtain the following subproblems for the primal iterates:

Primal updates:

$$
\begin{aligned}
\mathbf{v}^{k+1} &= \underset{\mathbf{v}=(v_1,v_2,v_3)}{\arg\min} \left\{ \frac{1}{2} \int_\Omega (\rho_t + \mathbf{v}\cdot\nabla\rho)^2 \;+\; \left\langle p_1^k, \nabla v_1 - z_1^k \right\rangle \;+\; \frac{\mu}{2} \left\| \nabla v_1 - z_1^k \right\|_2^2 \right. \\
&\qquad \left. +\; \left\langle p_2^k, \nabla v_2 - z_2^k \right\rangle \;+\; \frac{\mu}{2} \left\| \nabla v_2 - z_2^k \right\|_2^2 \;+\; \left\langle p_3^k, \nabla v_3 - z_3^k \right\rangle \;+\; \frac{\mu}{2} \left\| \nabla v_3 - z_3^k \right\|_2^2 \right\} \\
&= \underset{\mathbf{v}=(v_1,v_2,v_3)}{\arg\min} \left\{ \frac{1}{2} \int_\Omega (\rho_t + \mathbf{v}\cdot\nabla\rho)^2 \;+\; \frac{\mu}{2} \left\| z_1^k - \nabla v_1 - \frac{1}{\mu}\, p_1^k \right\|_2^2 \right. && \text{(I)} \\
&\qquad \left. +\; \frac{\mu}{2} \left\| z_2^k - \nabla v_2 - \frac{1}{\mu}\, p_2^k \right\|_2^2 \;+\; \frac{\mu}{2} \left\| z_3^k - \nabla v_3 - \frac{1}{\mu}\, p_3^k \right\|_2^2 \right\} \\
z_1^{k+1} &= \underset{z_1}{\arg\min} \left\{ \frac{\mu}{2} \left\| z_1 - \left(\nabla v_1^{k+1} + \frac{1}{\mu}\, p_1^k \right) \right\|_2^2 \right\} \;+\; \alpha \int_\Omega |z_1| && \text{(II)} \\
z_2^{k+1} &= \underset{z_2}{\arg\min} \left\{ \frac{\mu}{2} \left\| z_2 - \left(\nabla v_2^{k+1} + \frac{1}{\mu}\, p_2^k \right) \right\|_2^2 \right\} \;+\; \alpha \int_\Omega |z_2| && \text{(III)} \\
z_3^{k+1} &= \underset{z_3}{\arg\min} \left\{ \frac{\mu}{2} \left\| z_3 - \left(\nabla v_3^{k+1} + \frac{1}{\mu}\, p_3^k \right) \right\|_2^2 \right\} \;+\; \alpha \int_\Omega |z_3| \;, && \text{(IV)}
\end{aligned}
$$

where we mainly shifted the scalar products to the quadratic relaxation terms. Note that the z-updates are simple sparsity regularization functionals we introduced in Section 4.4.3. Thus, a solution can be computed explicitly and very efficiently via thresholding using the following soft shrinkage formula:

$$
z_i^{k+1} \;=\; S\left(\nabla v_i^{k+1} + \frac{1}{\mu}\, p_i^k \,,\; \frac{\alpha}{\mu} \right) \quad \text{, for } i = 1, .., 3 \;.
$$

For the computation of $\mathbf{v}$ we will apply a Jacobi iteration. The optimality system for (I) reads as follows

$$\left\{\begin{array}{rcl} 0 &=& \left(\rho_t + v^{k+1}\nabla\rho\right)\rho_{x_1} \;+\; \mu\,\nabla\cdot\left(z_1^k - \nabla v_1^{k+1} - \frac{1}{\mu}\,p_1^k\right) \\ 0 &=& \left(\rho_t + v^{k+1}\nabla\rho\right)\rho_{x_2} \;+\; \mu\,\nabla\cdot\left(z_2^k - \nabla v_2^{k+1} - \frac{1}{\mu}\,p_2^k\right) \\ 0 &=& \left(\rho_t + v^{k+1}\nabla\rho\right)\rho_{x_3} \;+\; \mu\,\nabla\cdot\left(z_3^k - \nabla v_3^{k+1} - \frac{1}{\mu}\,p_3^k\right) \end{array}\right. .$$

To derive a Jacobi iteration in this case, we have to select a discretization of the Laplacian of a scalar field u. Here we choose a 5 point star in 2D, respectively a 7 point star in 3D, i.e.

$$\begin{aligned} [\Delta u]_{i,j,k} &= u_{i+1,j,k} + u_{i,j+1,k} + u_{i,j,k+1} + u_{i-1,j,k} + u_{i,j-1,k} + u_{i,j,k-1} - 6\,u_{i,j,k} \\ &=: \bar{u} - 6\,u_{i,j,k} \;. \end{aligned}$$

Thus, we have to solve the following linear system of equations

$$\begin{pmatrix} \rho_{x_1}^2 + 6\mu & \rho_{x_1}\rho_{x_2} & \rho_{x_1}\rho_{x_3} \\ \rho_{x_1}\rho_{x_2} & \rho_{x_2}^2 + 6\mu & \rho_{x_2}\rho_{x_3} \\ \rho_{x_1}\rho_{x_3} & \rho_{x_2}\rho_{x_3} & \rho_{x_3}^2 + 6\mu \end{pmatrix} \begin{pmatrix} v_1^{k+1} \\ v_2^{k+1} \\ v_3^{k+1} \end{pmatrix} = \begin{pmatrix} -\rho_{x_1}\rho_t - \mu\nabla\cdot\left(z_1^k - \frac{1}{\mu}\,p_1^k\right) - \mu\,\bar{v_1}^k \\ -\rho_{x_2}\rho_t - \mu\nabla\cdot\left(z_2^k - \frac{1}{\mu}\,p_2^k\right) - \mu\,\bar{v_2}^k \\ -\rho_{x_3}\rho_t - \mu\nabla\cdot\left(z_3^k - \frac{1}{\mu}\,p_3^k\right) - \mu\,\bar{v_3}^k \end{pmatrix} .$$

Applying Cramer's rule we obtain the following formulas for the vector field components

$$v_1^{k+1} = \frac{det(A_1^k)}{det(A)}, \quad v_2^{k+1} = \frac{det(A_2^k)}{det(A)}, \quad v_3^{k+1} = \frac{det(A_3^k)}{det(A)} \;,$$

where A denotes the matrix above, and where A_i denotes A with the i-th column replaced by the right hand side. The determinant of A can be computed in advance, which improves efficiency significantly. The Jacobi iteration updating v_1, v_2 and v_3 in an alternating manner can be repeated several times. However, in most of our test cases we needed only one inner iteration. In the following section we will apply this algorithm to 2D and 3D synthetic and real life applications.

8.6 Numerical Results

8.6.1 Results in Computed Tomography

In this subsection we present results in Computed Tomography (CT) for heart structure analysis. The following structural motion results have been developed in collaboration with Paul Lunkenheimer from the university hospital in Münster. We

acknowledge the SCANCO Medical AG in Bruettisellen, Switzerland for providing medical imaging data.

The cardiac muscle, called myocardium is a type of involuntary striated muscle found in the walls of a heart. To analyze the structure of the cardiac muscle we investigated high resolution $3D$ data sets of a heart. The data size is $2304 \times 2304 \times 1336$.

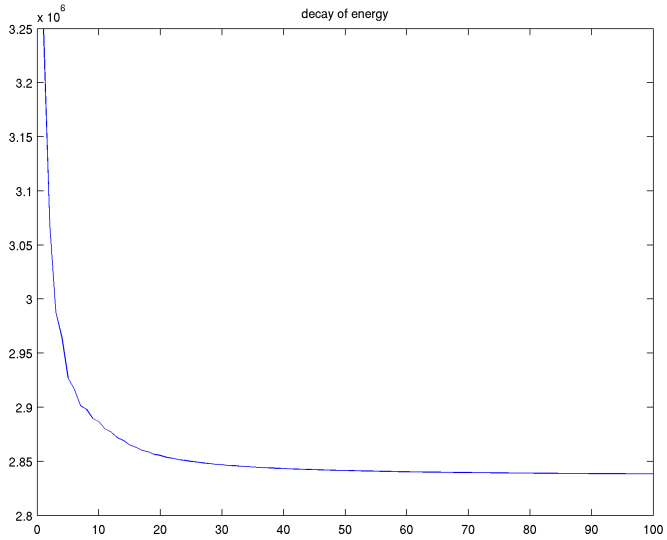


Figure 8.9: Decrease of functional values

8.7 Compressible versus Incompressible Flows

Similarly to decomposition models for images (scalar fields), e.g. cartoon-texture decomposition, there exists a (natural) decomposition of vector fields. More precisely, we have an orthogonal decomposition of $L^2(\Omega)^2$, a so-called Hodge-decomposition of $L^2(\Omega)^2$,

$$L^2(\Omega)^2 \;=\; \mathcal{H} \;\oplus\; \nabla H_0^1(\Omega) \;\oplus\; \nabla^\perp H_0^1(\Omega) \;,$$

where the gradient and curls of H^1 are characterized by

$$\begin{aligned} \nabla^\perp H^1(\Omega) \;&:=\; \{\mathbf{v} \in H(div;\Omega) \mid div(\mathbf{v}) = 0\} \\ \nabla H^1(\Omega) \;&:=\; \{\mathbf{v} \in H(curl;\Omega) \mid curl(\mathbf{v}) = 0\} \;, \end{aligned}$$

and where

$$\mathcal{H} \;:=\; \{\mathbf{v} \in H(div;\Omega) \cap H(curl;\Omega) \mid div(\mathbf{v}) = curl(\mathbf{v}) = 0\} \;,$$

denotes the space of harmonic vector fields, i.e. those with $\Delta \mathbf{v} = 0$. The Hodge decomposition results from two Helmholtz-decomposition of $L^2(\Omega)^2$. For further technical details we refer to [196, 195, 52]

In other words, we can characterize a vector field $\mathbf{v}$ by a sum of a divergence part, a curl part and harmonic part. The curl of a flow field describes the vorticity, whereas the divergence $\nabla \cdot \mathbf{v}$ is characterized by sources and sinks. For example take a look at the simple vector fields in Figure 8.3 at the beginning of this chapter in Section 8.1. The rotation example is divergence-free, the zoom-in is curl-free and the translation has no rotation and not sources or sinks.

In fluid mechanics, or more generally, continuum mechanics, an incompressible flow (more precisely an isochoric flow) is a solid or a fluid flow in which the divergence of the velocity is zero, i.e.

$$\nabla \cdot \mathbf{v} = 0 \ ,$$

i.e. one does not allow any source or sinks in the flow. At the beginning of this chapter we saw that the common brightness constancy assumption in optical flow leads to the optical flow constraint concerning $\mathbf{v}$, i.e.

$$\rho(x + \mathbf{v}, t + \delta t) \stackrel{!}{=} \rho(x, t) \tag{8.36}$$

$$\stackrel{Taylor}{\rightsquigarrow} \quad \partial_t \rho + \mathbf{v} \ \nabla \rho = 0 \ . \tag{8.37}$$

This is equivalent to saying that the material derivative $\frac{D}{Dt}$ of the density ρ is zero. Now we go over to a more general equation, the so-called continuity equation,

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = \partial_t \rho + \mathbf{v} \ \nabla \rho + \rho \ \nabla \cdot \mathbf{v} = 0 \quad \text{(mass conservation)} \ . \tag{8.38}$$

We will study this equation further in detail in the following chapter in the context of continuum mechanics. The Figure 9.2 in the following chapter illustrates the different view points of optical flow and mass conservation. The red-marked term in 8.38 is equal to the rate of brightness formation from the sources or sinks in an image flow. If we compare 8.38 and 8.37, the optical flow model assumes

$$\rho \nabla \cdot \mathbf{v} \ = \ 0 \ .$$

Hence, optical flow models are suitable for estimating incompressible flows, i.e. where the corresponding density does not change due to time. That is one reason why optical flow is useful for tracking problems, e.g. for car traffic scenes.

However, for instance in cardiac gate positron emission tomography (PET) in nuclear medicine, we can observe that densities can change intensively due to time. For example take a look at the two time steps of a pumping heart sequence in Figure 8.22.

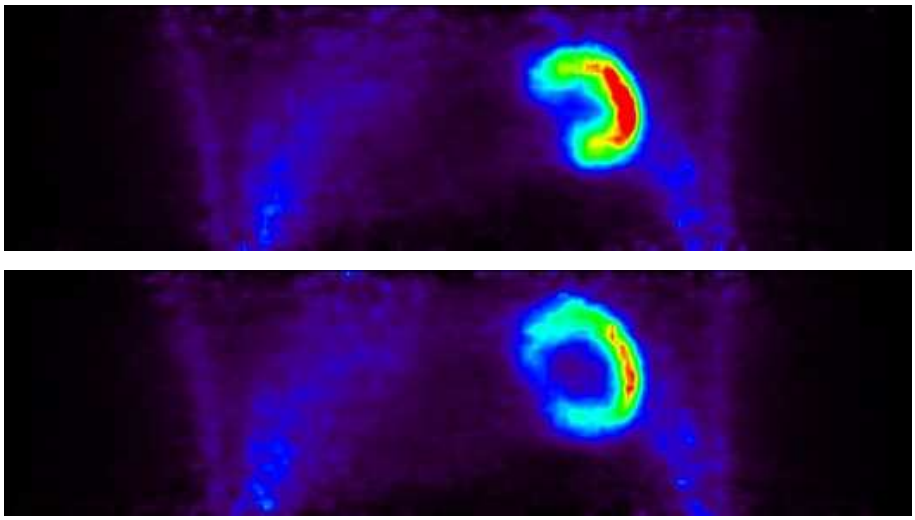


Figure 8.22: Two time steps of a cardiac gated PET data set; conservation of mass concerning the injected tracer, but large density changes due to time

Nevertheless, we can expect a conservation of mass due to time concerning an injected tracer. The continuity equation from above 8.38 allows compressible flows, i.e. it allows density changes of investigated objects due to time, whereas the mass is preserved. As a consequence, mass conservation is a reasonable motion model for PET images, in particular for cardiac data sets. In the following chapter we will see that the continuity equation can be seen as a fundamental part of optimal transport models. The allowance of compressible flows and the relation to continuum mechanics are reasons why we decided to combine $4D$ image reconstruction with optimal transport models in Chapter 10. Now we proceed with continuum mechanics, conservation laws and optimal transport.

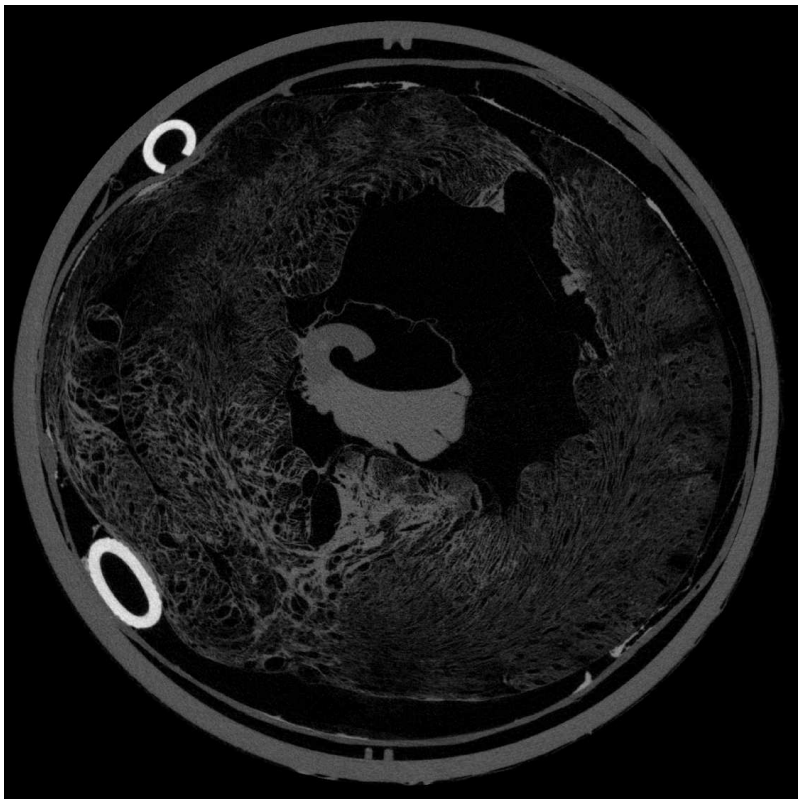


Figure 8.11: Slice 633 of 1336 in z-direction

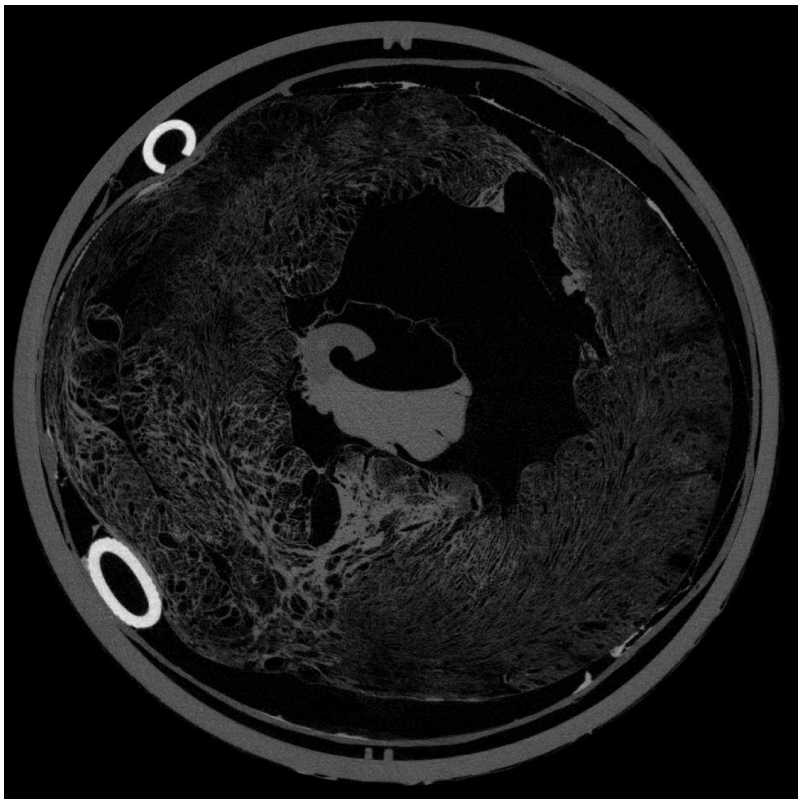


Figure 8.12: Slice 634 of 1336 in z-direction

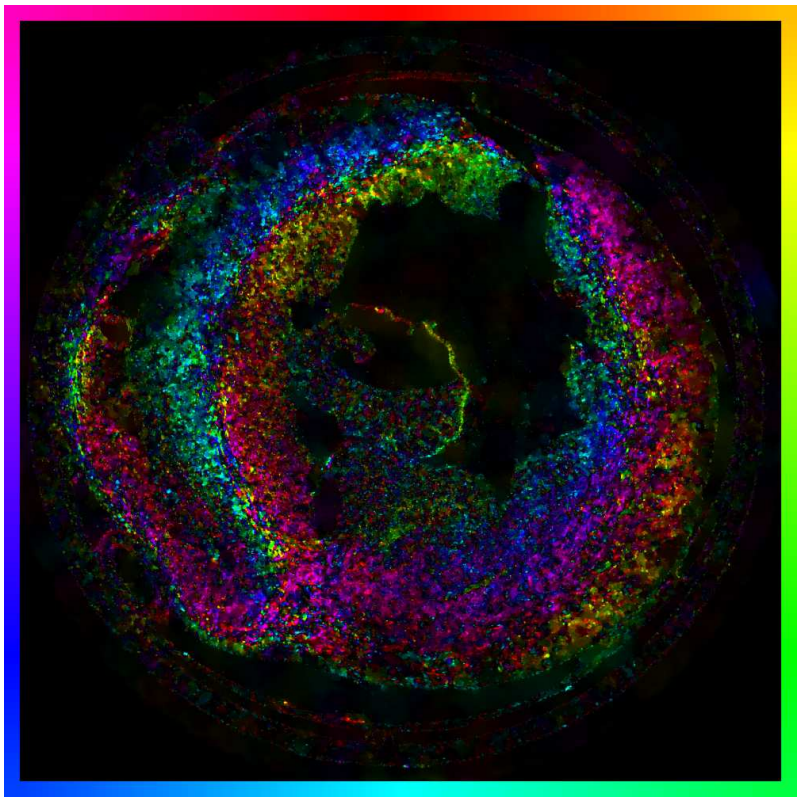


Figure 8.13: Optical flow $\mathbf{v}$ in colorful orientation plot between slice 633 and 634

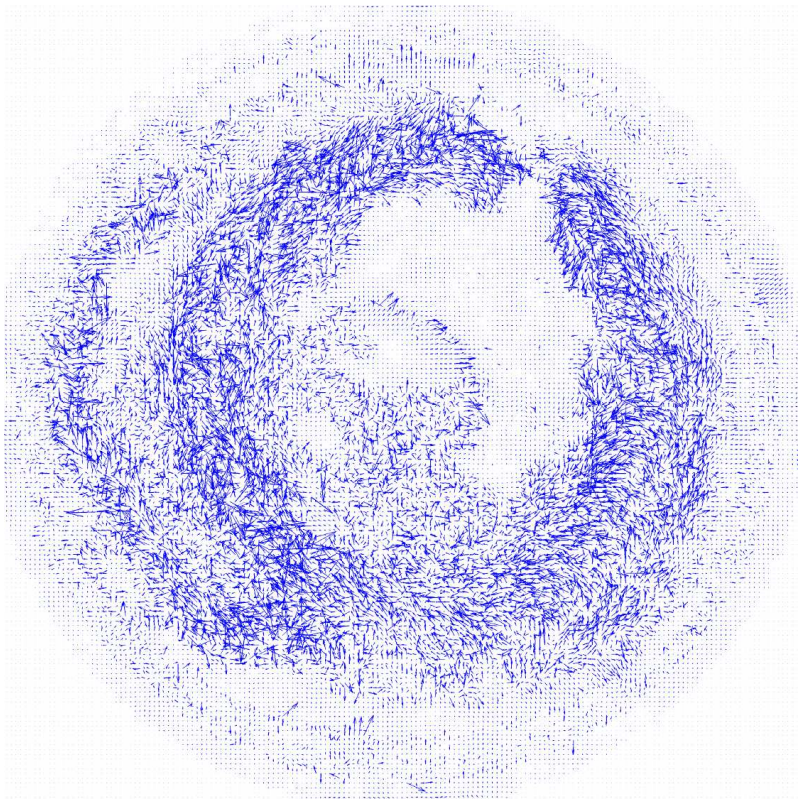


Figure 8.14: Optical flow $\mathbf{v}$ in vector plot between slice 633 and 634

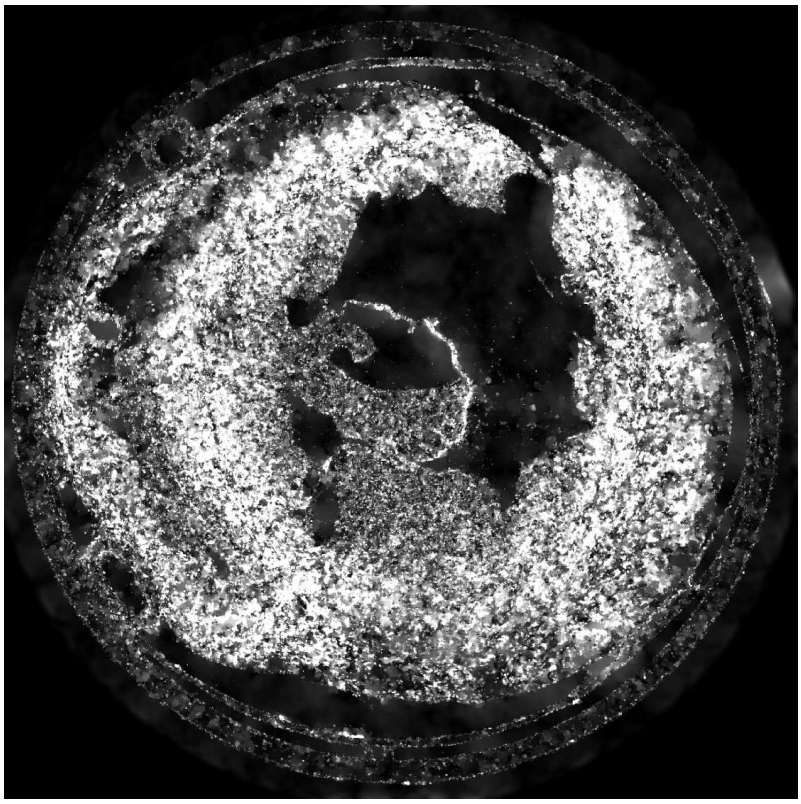


Figure 8.15: Magnitude of the optical flow $\mathbf{v}$ between slice 633 and 634

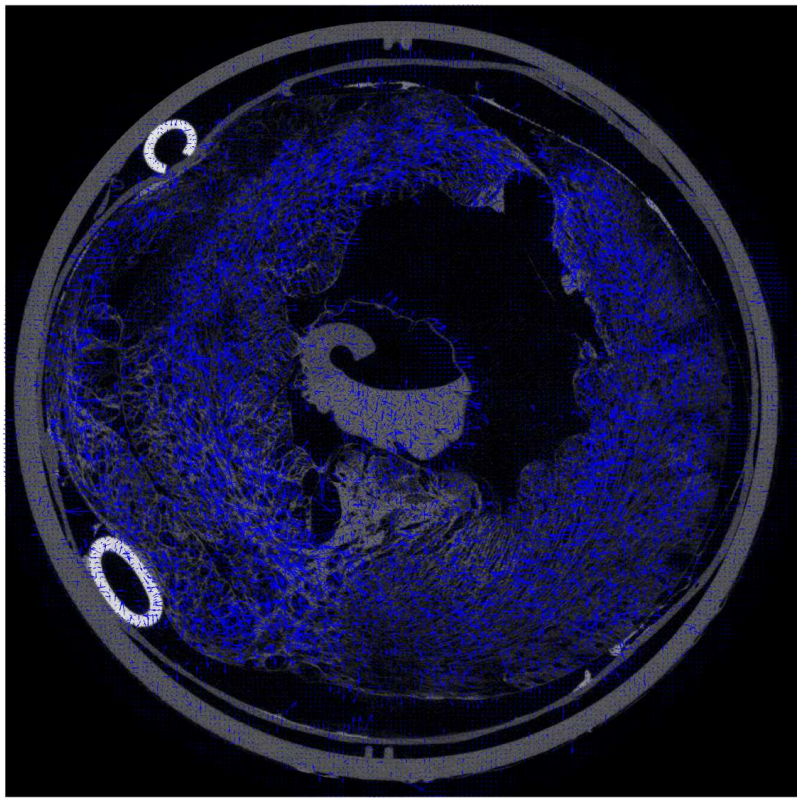


Figure 8.16: Slice 633 and overlay of vector field $\mathbf{v}$ between slice 633 and 634

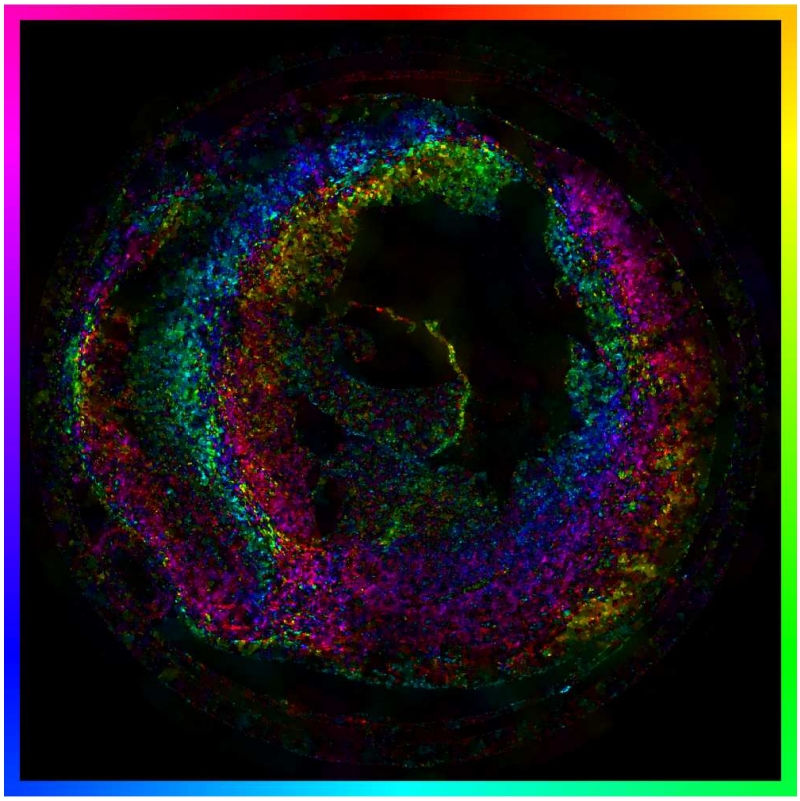


Figure 8.17: Optical flow $\mathbf{v}$ in colorful orientation plot between slice 634 and 635

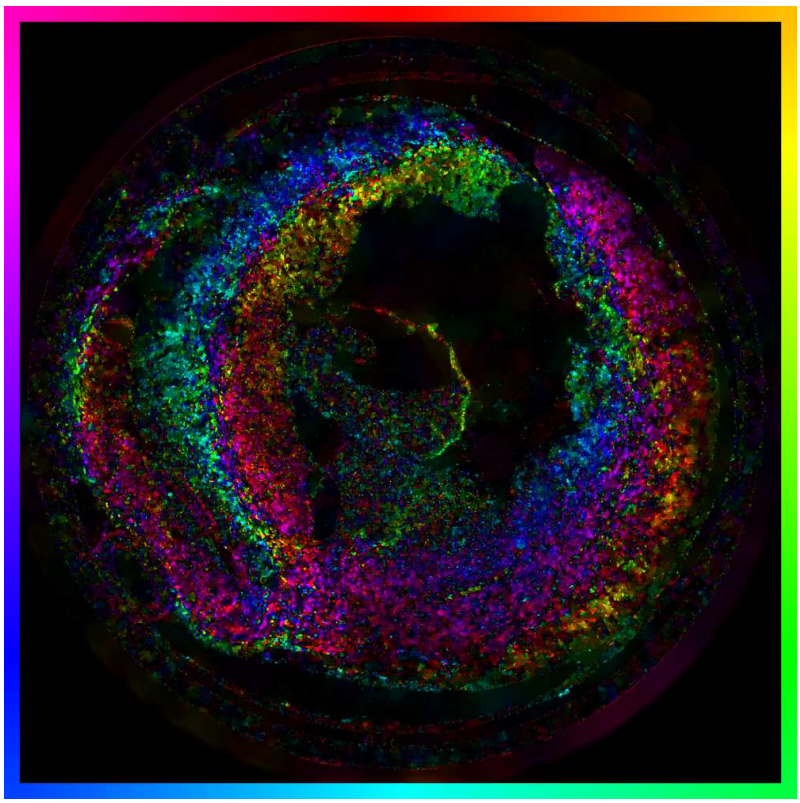


Figure 8.18: Optical flow $\mathbf{v}$ in colorful orientation plot between slice 635 and 636

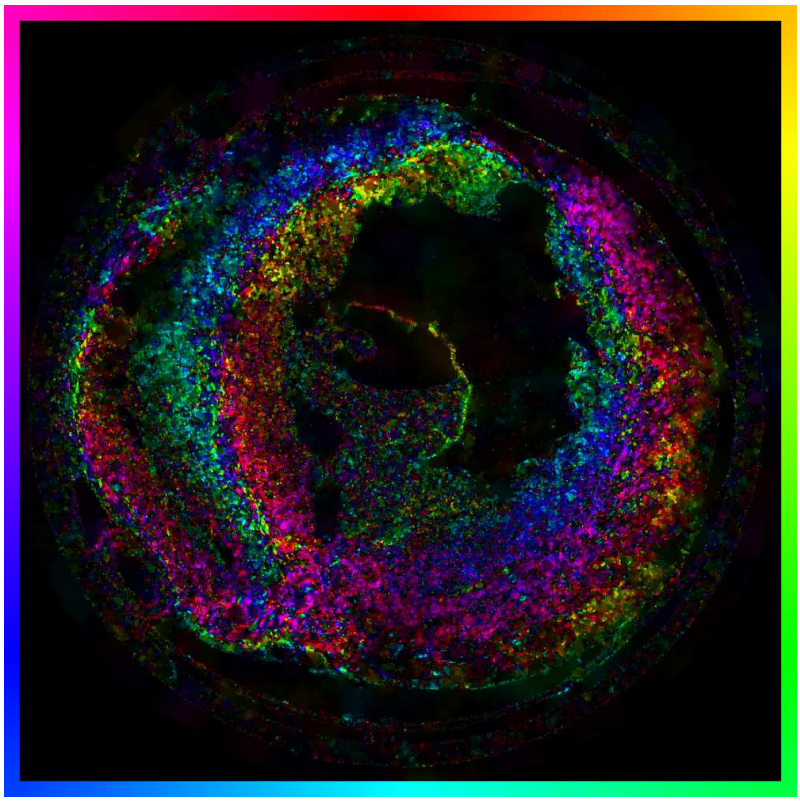


Figure 8.19: Optical flow $\mathbf{v}$ in colorful orientation plot between slice 636 and 637

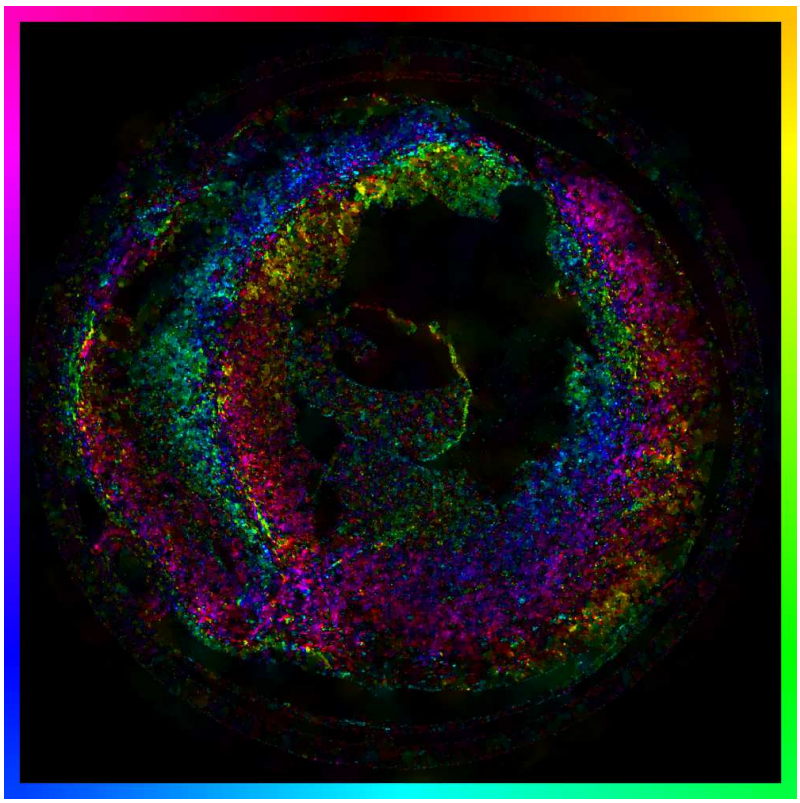


Figure 8.20: Optical flow $\mathbf{v}$ in colorful orientation plot between slice 637 and 638

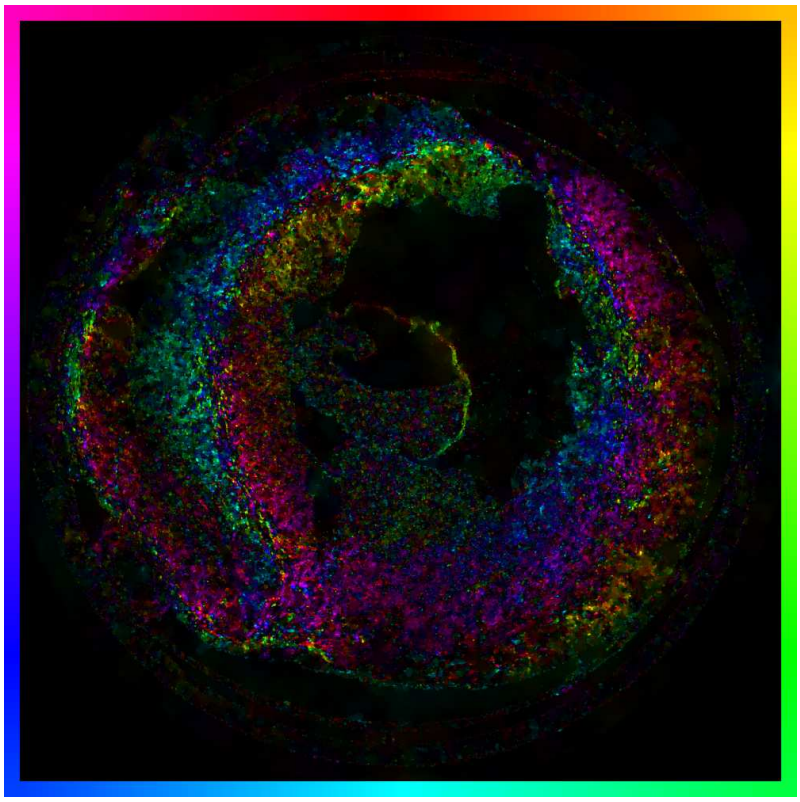


Figure 8.21: Optical flow $\mathbf{v}$ in colorful orientation plot between slice 638 and 639

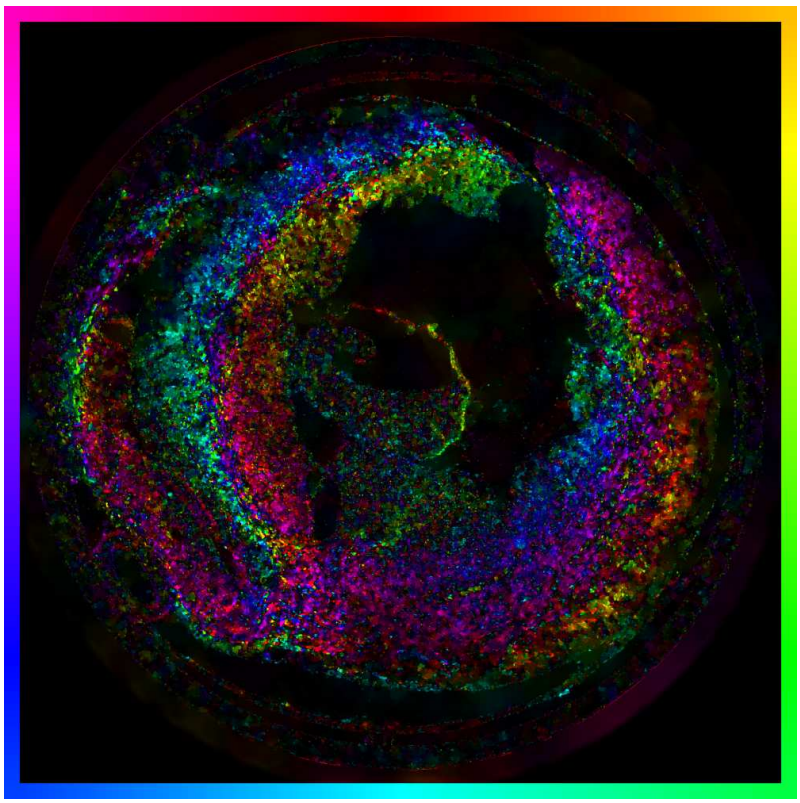


Figure 8.22: Optical flow $\mathbf{v}$ in colorful orientation plot between slice 639 and 640

9 Optimal Transport

In this chapter we will introduce a fundamental equation of continuum mechanics to motivate optimal transport. Then we will relate optimal transport to optical flow estimation from the previous chapter. In the following chapters optimal transport with mass conservation will serve as a conceptional ingredient of 4D imaging.

9.1 Continuum mechanics

Continuum mechanics is a branch of mechanics that deals with the analysis of the kinematics and the mechanical behavior of materials modeled as a continuum. Modeling an object as a continuum assumes that the substance of the object completely fills the space it occupies. In the following we will concentrate on fluid mechanics, a branch of continuum mechanics. For detailed models of solid mechanics, which is an other branch of continuum mechanics, we refer e.g. to [174].

Fluid mechanics is the study of fluids and the forces on them. Fluid flows can be seen as physical mass continua, in other words, objects in the Euclidean space which can be seen as the set of their particles or mass points. The derivation of the equations of fluid dynamics are based on the following main physical principles:

- For all times $t > 0$, there exists a well-defined mass density $\rho(x, t)$, such that the mass $m(\Omega, t)$ in the region Ω at time t is given by

$$m(\Omega, t) = \int_{\Omega} \rho(x, t) dx \tag{9.1}$$

- Mass is neither created nor destroyed. (*mass conservation*)

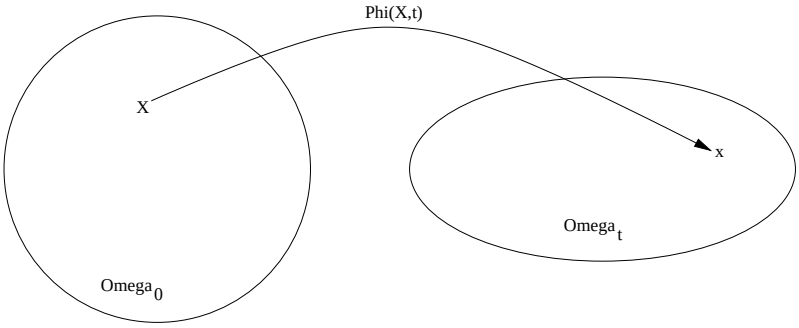
- The changes of the momentum in a fluid region equals the attached forces. (*momentum conservation*)
- Energy is neither created nor destroyed. (*energy conservation*)

These assumptions are called continuum hypothesis. These foundational axioms of fluid dynamics are the conservation laws, specifically, *conservation of mass*, *conservation of linear momentum* (also known as Newton's second law of motion), and *conservation of energy* (also known as first law of thermodynamics).

Let $\Omega \subset \mathbb{R}^d$ be the region occupied by the fluid. Let $x \in \Omega$ and consider the fluid particle X, which moves at time t though position x. For describing the mass continuum, x are called spatial or *Euler coordinates*, whereas X are termed material or *Lagrangian coordinates*. Let $W_0 \subset \Omega$ be a subset at time $t = 0$. The function $\phi : W \times \mathbb{R}^+ \to \mathbb{R}^d$ describes the change of the particle position

$$W_t := \{\phi(X,t) : X \in W_0\} = \phi(W_0, t) \ .$$

The *pathline* is the set of space points $x(X_0, t)$ that is taken by a particle X_0 at



different times t. In other words, pathlines are the trajectories that fluid particles follow.

A *streamline* is a curve that is instantaneously tangent to the velocity vector of the flow. This shows the direction a fluid element will travel in at any point in time (see Fig. 9.1). Streamlines represent contours of the velocity field showing the motion of the whole field at the same time. In the case of stationary flow fields, that is

Streaklines are the locus of points of all the fluid particles that have passed continuously through a particular spatial point in the past. For example a dye steadily injected into a fluid at a fixed point extends along a streakline. Appropriate to the previous section on optical flow the velocity of a particle will be denoted by $u(x,t)$ in the following. For fixed time t, $u(x,t)$ is a vector field on Ω. Then

$$\begin{aligned} x : \ \mathbb{R}^+ &\to \mathbb{R}^d \\ t &\mapsto \phi(X,t) \end{aligned}$$

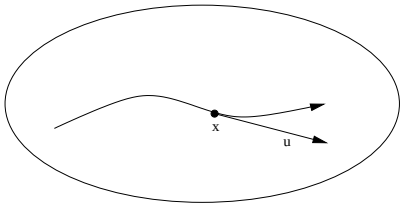


Figure 9.1: A streamline and the flow field u

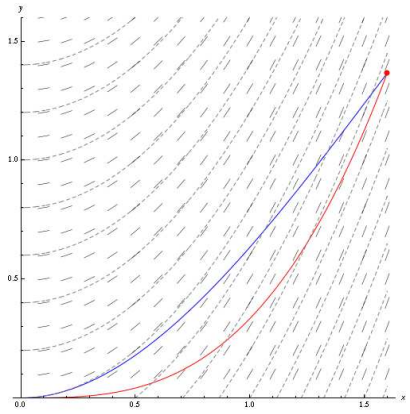


Figure 9.2: Pathline(red dot), streamlines(dashed lines) and streaklines(blue dot)

is the particle path line and the velocity field u is given by

$$u(x,t) = \frac{\partial \phi}{\partial t}(X,t), \qquad \text{with } x = \phi(X,t) \ .$$

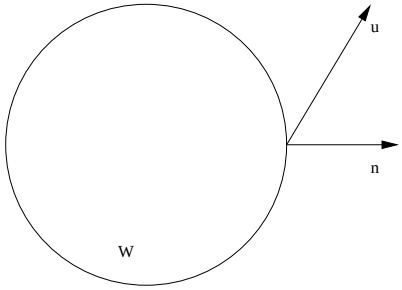
The acceleration of a particle is obtained from deriving the velocity field and by use of the chain rule

$$\begin{aligned}
\frac{d}{dt}u(x,t) &= \frac{d}{dt}u(\phi(X,t),t) \\
&= \frac{\partial}{\partial t}u(\phi(X,t),t) + \sum_{i=1}^{d} \frac{\partial u}{\partial x_i}(\phi(X,t),t) \underbrace{\frac{\partial \phi_i}{\partial t}(X,t)}_{=u_i(x,t)} \\
&= \frac{\partial}{\partial t}u(x,t) + \sum_{i=1}^{d} u_i(x,t)\frac{\partial u}{\partial x_i}(x,t) \\
&= \frac{\partial u}{\partial t} + (u \cdot \nabla)u \ .
\end{aligned}$$

The resulting operator

$$\frac{D}{Dt} \ := \ \frac{\partial}{\partial t} + u \cdot \nabla$$

is known as the *material derivative* towards t.



9.2 Mass conservation

In this section we derive a motion equation from the mass conservation assumption in the continuum hypothesis from above. Therefore we fix a region $W \subset \Omega$. ∂W denotes the boundary of W, let n be an outer normal vector and dS describes the area element on $\partial\Omega$. Assuming differentiability of the integrand and integrability of the derivative, the variation of mass in W is given by

$$\frac{d}{dt} m(W,t) \overset{(9.1)}{=} \frac{d}{dt} \int_W \rho(x,t)\, dx \tag{9.2}$$

$$= \int_W \frac{\partial}{\partial t} \rho(x,t)\, dx\ . \tag{9.3}$$

As u and n are perpendicular the volumetric flow rate towards ∂W equals $u \cdot n$. It is the volume of fluid that passes through a unit area of the boundary per unit time. The corresponding mass flow rate is $\rho u \cdot n$. Hence the mass flow rate towards ∂W is given by

$$\int_{\partial W} \rho\, u \cdot n\, dS\ .$$

True to the motto "Nothing can come of nothing" the principle of mass conservation means that a variation of mass in W equals the mass flow rate towards the boundary ∂W (introversive), more precisely

$$\frac{d}{dt} \int_W \rho\, dV \overset{!}{=} -\int_{\partial W} \rho\, u \cdot n\, dS \overset{Gauss}{=} -\int_W u\, \nabla\rho - \int_W \rho\, \nabla \cdot u\ .$$

Hence equation (9.3) and Gauss's integral theorem imply

$$\int_W \left(\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) \right) dx = 0\ .$$

Since we started with an arbitrary subset W, we can conclude the PDE form of the *continuity equation*.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0\ .$$

Alternatively, this conservation law can be obtained from discrete particle mechanics models by passing over to the continuum, i.e. number of mass points $N \to \infty$.

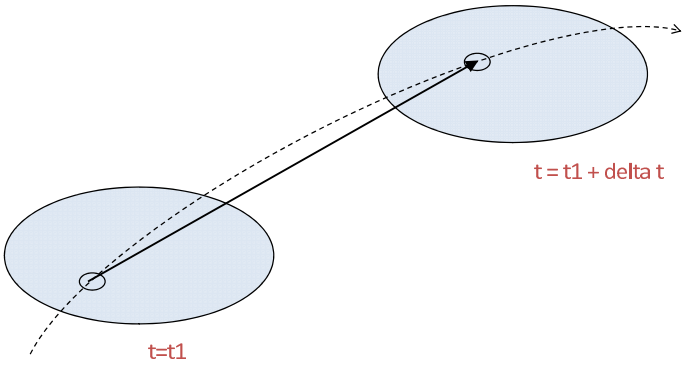


Figure 9.4: The optical flow approach

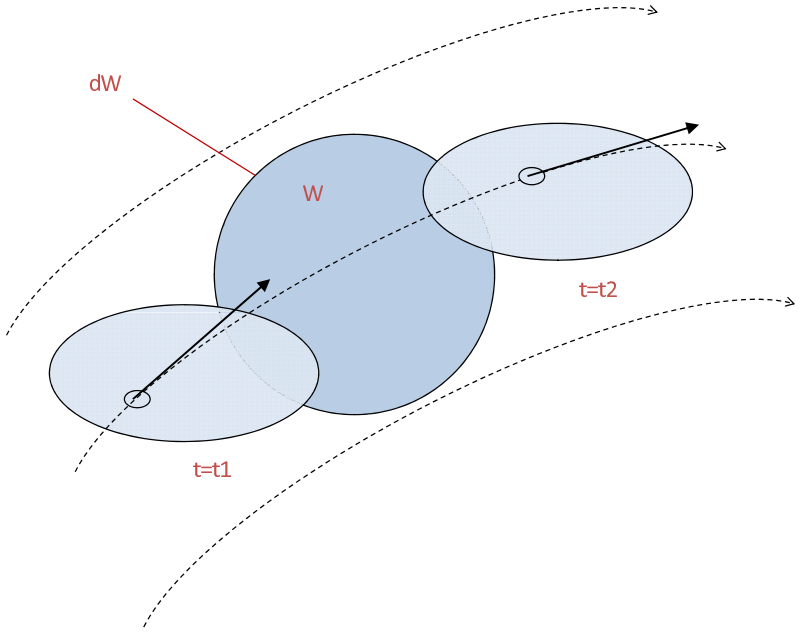
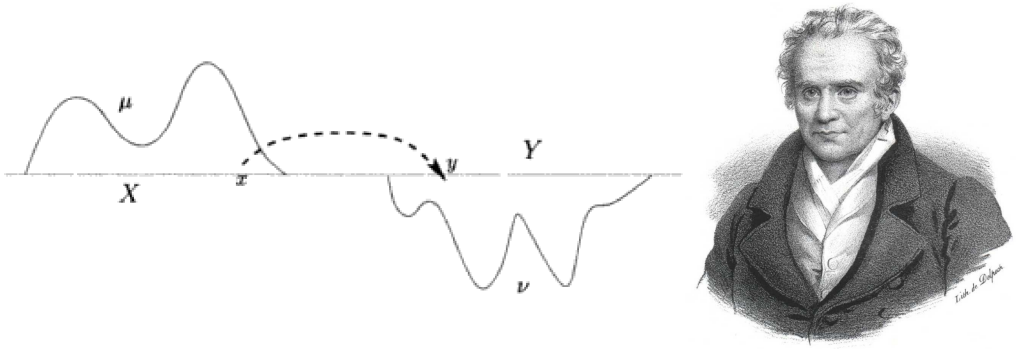


Figure 9.5: The transport theory approach

9.3 Monge´s Transport Problem

The original transport problem can be traced back to a work of the French mathematician Gaspard Monge published in 1781 [128].



The basic idea of his problem was to find a minimal-cost transport strategy for moving a heap of sand into a hole of the same size. To give a mathematical description of the Monge problem we first of all have to introduce pushforward operators .

Definition 9.3.1 (Pushforward operator)**.** *Given measurable space* (X, d) *and* (Y, d)*, a measurable function* $r : X \to Y$ *be an injective transformation. The pushforward of* μ *through* r

$$r_*(B) := \mu(r^{-1}(B)) = \nu(B), \qquad \forall \text{ measurable } B \subset Y$$

For given μ and ν the Monge problem reads as follows

$$\inf_r \left\{ \int_X d(x, r(x))\, d\mu(x) \mid r_*\mu = \nu \right\} .$$

The interpretation of this minimization problem is to find a transformation r under the push forward constraint such that the transportation cost is minimal. The problem of this formulation is the fact that neither existence nor uniqueness of a solution can be expected. Simple examples with Dirac delta distributions show that the pushforward operator does not allow to "split" densities. To overcome this problem Kantorovich suggested a relaxed formulation of the model,

$$\frac{1}{2} \int_{\Omega\times\Omega} |x - y|^2 d\Pi(x, y) + \epsilon E(u) \to \min_{\Pi,u},$$

subject to

$$\begin{aligned} \int_{A\times\Omega} d\Pi(x, y) &= \int_A d\nu(y) \\ \int_{\Omega\times A} d\Pi(x, y) &= \int_A u(x)dx \end{aligned}$$

for all $A \subset \Omega$ measurable, where u is a probability density in the domain of E and Π is a probability measure on $\Omega \times \Omega$.

The Wasserstein metric is a distance function defined between probability distributions on a given metric space. If we interpret each distribution as a unit mass piled on a domain X, the metric is the cost of turning one pile into the other. This cost is assumed to be the amount of mass that needs to be moved times the distance it has to be moved. Due to this analogy the metric is sometimes called *earth mover's distance*.

Definition 9.3.2 (p-Kantorovich-Wasserstein distance)**.** *Let (X, d) be a separable metric space and $\mu \in \mathbb{P}(X)$ a probability measure on X. Let further X satisfy the Radon property, i.e.,*

$$\forall B \in \mathbb{B}(X), \epsilon > 0 \text{ there exists an } K_\epsilon \subset\subset B : \quad \mu(B \setminus K_\epsilon) \leq \epsilon,$$

and let $p \geq 1$. The $(p-th)$ Wasserstein distance between two probability measures $\mu^1, \mu^2 \in \mathbb{P}_p(X)$ (the space of all probability measures on X with μ-integrable $p-th$ moments) is defined by

$$W_p(\mu^1, \mu^2)^p := \min_{\Pi \in \Gamma(\mu^1, \mu^2)} \int_{X^2} d(x, y)^p \, d\Pi(x, y).$$

Here $\Gamma(\mu^1, \mu^2)$ denotes the class of all transport maps $\gamma \in \mathbb{P}(X^2)$ such that

$$\pi^1_\# \gamma = \mu^1, \quad \pi^2_\# \gamma = \mu^2,$$

where $\pi^i : X^2 \to X$, $i = 1, 2$, are the canonical projections on X, and $\pi^i_\# \gamma \in \mathbb{P}(X)$ is the push-forward of γ through π^i.

In our setting, i.e. in the case of transporting $2D$ oder $3D$ density images, $\Omega \subset \mathbb{R}^d$ is an open and bounded domain and $|\cdot|$ is the usual Euclidean distance on $\mathbb{R}^d$. Hence, the assumptions for the Wasserstein distance above are fulfilled.

I will shortly summarize the Monge-Kantorovich framework for our needs in imaging: For two positive densities $\rho_0(x) \geq 0$ and $\rho_T(x) \geq 0$ with $x \in \Omega \subset \mathbb{R}^d$, which have a total mass bounded by one, i.e.

$$\int_{\mathbb{R}^d} \rho_0(x) \, dx = \int_{\mathbb{R}^d} \rho_T(x) \, dx = 1 \ ,$$

we are interested in computing the L^p-Kantorovich-Wasserstein distance

$$W_p(\rho_0, \rho_T)^p \ := \ \inf_M \int |M(x) - x|^p \, \rho_0(x) \, dx \ .$$

9.4 The Benamou-Brenier Formulation

To compute the L^2-Kantorovich-Wasserstein distance, Benamou and Brenier [17, 18] suggested an equivalent formulation in the year 2000. They transferred the problem into a continuum mechanics framework which we have introduced in the first section of this chapter to some extent.

Benamou and Brenier fixed an time interval $[0, T]$ and assumed two given densities ρ_0 and ρ_T as initial and final condition. They considered all possible, sufficiently smooth, time dependent densities and velocity fields, $\rho(x,t) \geq 0$, $\mathbf{v}(x,t) \in \mathbb{R}^d$, subject to the continuity equation for $0 < t < T$ and $x \in \Omega \subset \mathbb{R}^d$. Benamou and Brenier proved that computing the L^2-Wasserstein distance $W_2(\rho_0, \rho_T)^2$ is equivalent to solving the following constraint optimization problem

$$\inf_{\rho,\mathbf{v}} \; T \int_0^T \int_\Omega \rho(x,t) \; |\mathbf{v}(x,t)|^2 \, dx \, dt$$

$$\begin{aligned} \text{subject to} \qquad \partial_t \rho + \nabla_\Omega \cdot (\rho \, \mathbf{v}) &= 0 \qquad \text{(mass conservation)} \\ \rho(\cdot, 0) &= \rho_0 \\ \rho(\cdot, T) &= \rho_T \; . \end{aligned}$$

This transport model is based on mass conservation as a motion model. The objective functional couples the density ρ and the velocity field $\mathbf{v}$, and minimizes the cost of a transport plan. The result of this continuum mechanics formulation is a time interpolant $\rho(x,t)$ of the given data ρ_0 and ρ_T, and a velocity field $\mathbf{v}(x,t) \in \mathbb{R}^d$ which moves ρ_0 toward ρ_T.

It is interesting to see that a continuum mechanics formulation was already implicitly contained in the original work of Monge we described in the previous section. In his original work [128] eliminating the time variable was just a clever way of reducing the dimension of the problem. In physics the product of density and velocity $\mathbf{j} := \rho\mathbf{v}$ is well known as the momentum. As mentioned in the continuum hypothesis in Section 9.1, the momentum is in general a conserved quantity, too. This means, the total momentum of any closed system (one not affected by external forces) cannot change.

From a computational point of view the introduction of the momentum allows to solve a *convex* (although not quadratic) space-time minimization problem in the variables ρ and $\mathbf{j}$,

$$\inf_{\rho,\mathbf{j}} \; T \int_0^T \int_\Omega \frac{|\mathbf{j}(x,t)|^2}{\rho(x,t)} \, dx \, dt$$

$$\begin{aligned} \text{subject to} \qquad \partial_t \rho + \nabla_\Omega \cdot \mathbf{j} &= 0 \qquad \text{(mass conservation)} \\ \rho(\cdot, 0) &= \rho_0 \\ \rho(\cdot, T) &= \rho_T \ . \end{aligned}$$

The convexity results from the convexity of quadratic-over-linear functions, see Section 3.2, and the linearity of the constraints. Benamou and Brenier solved the space-time minimization problem above as a saddle point problem for a suitable Lagrangian. Recently, for image registration purposes, sophisticated numerical realizations of the L^2 optimal mass transfer problem have been proposed e.g. by [92]. An alternative to the Benamou-Brenier approach is to solve the Monge-Ampere partial differential equation,

$$det(\nabla M(x))\, \rho_T(M(x)) \ = \ \rho_0(x) \ .$$

directly. To gain a deeper insight into optimal transport models with more general measures and related analysis we refer to [3]. The somehow artificial interpolation time, which has been introduced for this continuum mechanics formulation, will be very useful for our 4D image reconstruction algorithms in the next chapter. In our practical applications in computer vision we will identify this (artificial) time with the "real" time in a given video sequence of densities. Hence, our motion model is motivated by optimal transport but is different in the sense that usually no initial or final densities are given.

10

4D Image Reconstruction in Nanoscopy and PET

In this section we will present 4D imaging models, more precisely, joint models including $4D$ image reconstruction, space-time regularization, and optimal transport techniques. Our models will reconstruct space-time densities in the sense of regularized inverse problems and simultaneously compute motion information regarding mass conservation. Particularly, in the interesting case of space-time TV density regularization we will provide a detailed analysis of our model including existence and uniqueness of a solution.

Moreover, this chapter deals with numerical methods for these joint $4D$ reconstruction models. The first numerical approach deals with the case of L^2 regularization. Here we follow the concept of "*first discretize and then optimize*" and derive Newton-SQP methods with line-search strategies and multigrid preconditioning. The second numerical approach follows the concept "*first optimize and then discretize*" and is strongly connected to convex splitting methods and fast Bregman techniques, see Chapter 4. These splitting strategies are based on inexact Uzawa type methods for the augmented Lagrangian. In the fashion of algorithms like Split Bregman or Bregmanized Operator Splitting, we apply suitable decoupling and preconditioning concepts. As as result, we obtain splitting algorithms for our 4D-TV model, where all subproblems are solvable very efficiently by FFT, respectively DCT or shrinkage and main parts of the implementation can be parallelized on GPUs.

10.1 $4D$ Model - Reconstruction and Optimal Transport

Inspired by commonly used $3D$ image reconstruction models and the optimal transport formulation of Benamou and Brenier [17] presented in Section 7.2 we present

the following constrained minimization problem. As an example, let us assume we have a given noisy and blurred image sequence in the sense of a time dependent inverse problem,

$$f(\cdot,t) = K\rho(\cdot,t) + \text{"noise"} \ . \tag{10.1}$$

Our new $4D$ imaging model reconstructs the density image sequence ρ from given data f in the sense of regularized inverse problems. Simultaneously, the underlying motion field $\mathbf{v}$ of the density is computed regarding optimal transport. For parameters $\alpha, \beta \geq 0$ our general model reads as follows

Model 10.1.1. *General 4D Reconstruction & Optimal Transport*

$$\begin{aligned} \min_{\rho,\, \boldsymbol{v}} \ & \int_0^T H_{f(\cdot,t)}\left(K\rho(\cdot,t)\right)\, dt \ + \ \alpha \int_0^T J(\rho(\cdot,t))\, dt \ + \ \frac{\beta}{2} \int_0^T \int_\Omega \rho\, |\boldsymbol{v}|^2 \, dx\, dt \\ s.t. \ & \partial_t\, \rho \ + \ \nabla_\Omega \cdot (\rho\, \boldsymbol{v}) \ = \ 0 \qquad \text{in } Q := \Omega \times [0,T] \\ & \rho \geq 0 \ . \end{aligned} \tag{10.2}$$

The first term of the objective functional is a space-time data fidelity term H, which depends on the given image sequence f, on a compact operator K and on the desired density sequence ρ. For simplicity, the compact operator K is assumed to have the same structure in all time steps t. In analogy to static reconstruction, different data fidelities result from different noise models using Bayes theory and MAP estimation. Hence in the cases of Gaussian noise and Poisson noise we get a L^2 norm, respectively a Kullback-Leibler data fidelity at each time step t in our model, i.e.

$$H_{f(\cdot,t)}\left(K\rho(\cdot,t)\right) \ := \ \frac{1}{2}\, \|K\rho(\cdot,t) - f(\cdot,t)\|^2_{L^2(\Sigma)} \qquad \text{(Gaussian noise)} \tag{10.3}$$

$$H_{f(\cdot,t)}\left(K\rho(\cdot,t)\right) \ := \ \int_\Sigma K\rho(\cdot,t) - f(\cdot,t)\log(K\rho(\cdot,t)) \qquad \text{(Poisson noise)} \ . \tag{10.4}$$

The second term of the objective functional in Model 10.2 is a time dependent regularization term. In the simple case of quadratic regularization we define J as

$$J(\rho(\cdot,t)) \ := \ \frac{1}{2}\, \|\nabla\rho(\cdot,t)\|_2^2 \ . \tag{10.5}$$

In the interesting case of space-time TV regularization we will choose J as

$$J(\rho(\cdot,t)) \ := \ \frac{1}{p}(\ |\rho(\cdot,t)|_{BV(\Omega)}\ )^p \ . \tag{10.6}$$

The third term in the objective functional and the mass conservation constraint build an incorporated optimal transport scheme. The third term is a coupling term that relates the density to the magnitude of the flow field in space-time. The term

can be interpreted as a weighted space-time regularization of the velocity field. The continuity equation is used to model the motion of the reconstructed density due to time. In comparison to an optical flow constraint as a motion model, the mass conservation is more general and also allows compressible flows. From an analytical point of view our 4D model is different from optimal transport, since usually no initial and final densities are given. Our density in the mass conservation is simultaneously "generated" via the data fidelity. With this in mind, our model is more related to analysis in mean field games. Furthermore, we lay emphasis on the positivity constraint. In optimal transport one usually starts with nonnegative initial values and consequently obtains nonnegative densities for all times from the transport model. Here we we need to enforce nonnegativity of the whole density in numerical algorithms.

10.2 $4D$ Model - Space-Time L^2 Regularization

Here we start with our joint model in the case of the L^2 data fidelity, see (10.3) and a quadratic regularization, see (10.5),

Model 10.2.1. *Joint 4D Model with L^2 Regularization*

$$\begin{aligned} \min_{\rho,\, \boldsymbol{v}} \; & \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Sigma)}\, dt \;+\; \frac{\alpha}{2}\int_0^T \|\nabla\rho\|_2^2 \; dt \;+\; \frac{\beta}{2}\int_0^T \int_\Omega \rho\, |\boldsymbol{v}|^2 \, dx\, dt \\ s.t. \quad & \partial_t\, \rho \;+\; \nabla_\Omega \cdot (\rho\, \boldsymbol{v}) \;=\; 0 \qquad in\ Q := \Omega \times [0,T] \\ & \rho \geq 0 \; . \end{aligned}$$

This model computes a reconstructed sequence of density images and simultaneously the motion in the whole sequence. The first term of the objective functional is a space-time data fidelity for the inverse problem in 10.1 concerning an additive Gaussian noise modeling. The term that couples ρ and $\mathbf{v}$ and the continuity equation represent the optimal transport scheme. The quadratic regularization term favors smooth density images in each time step in a spatial-homogeneous sense, i.e. for instance edges in the density images are not taken into account in a specific way.
In general, this constrained optimization problem is not convex. Hence, from a numerical and analytical point of view, it is reasonable to reformulate the model via the momentum substitution $\mathbf{j} := \rho\mathbf{v}$. In physics the momentum is a well-known term, particularly in continuum mechanics, see Chapter 9. Since we simply multiply $\mathbf{v}$ with a scalar field ρ, the momentum $\mathbf{j}$ has the same direction as the motion field

and we can easily resubstitute after computing a solution. With this substitution the system reads as follows

Model 10.2.2. *Joint 4D Model with L^2 Regularization*

$$\begin{aligned} \min_{\rho,\, \boldsymbol{j}} \; & \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Sigma)}\, dt \;+\; \frac{\alpha}{2}\int_0^T \|\nabla\rho\|_2^2 \; dt \;+\; \frac{\beta}{2}\int_0^T \int_\Omega \frac{|\,\boldsymbol{j}\,|^2}{\rho}\, dx\, dt \\ s.t. \quad & \partial_t\, \rho \;+\; \nabla_\Omega \cdot \boldsymbol{j} \;=\; 0 \qquad \textit{in } Q := \Omega \times [0,T] \\ & \rho \geq 0 \; . \end{aligned}$$

This problem is convex. The objective functional is convex, since the coupling term is a quadratic-over-linear function, see 3.1.1, and the constraints are convex since the mass conservation constraint is linear in ρ, respectively $\mathbf{j}$. This problem will be solved numerically via a Newton-SQP algorithm in Section 10.4.

Existence and Uniqueness
With the quadratic regularization term we have an a-priori estimate for

$$\rho \;\text{ in } L^p(0,T;W^{1,2}(\Omega)) \; ,$$

and with an additional substitution $\mathbf{w} := \sqrt{\rho}\mathbf{v}$ we obtain an a priori estimate for

$$w \;\text{ in } L^2((0,T)\times\Omega) \; .$$

Thus, following the proofs in the next section on TV regularization and replacing BV by $W^{1,2}(\Omega)$, the same arguments may yield existence and uniqueness of a minimizer.

10.3 $4D$ Model - Space-Time TV regularization

Inspired by the optimal transport formulation of Benamou and Brenier [17] presented in Section 7.2 and commonly used image reconstruction models based on total variation regularization (compare with the static reconstruction models in Section 5), we consider the following constrained minimization problem,

Model 10.3.1. *Joint 4D Model with Space-Time TV Regularization*

$$\begin{aligned} \min_{\rho,\, \boldsymbol{v}} \; & \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Sigma)}\, dt \;+\; \frac{\alpha}{p}\int_0^T (\,|\rho|_{BV(\Omega)}\,)^p \; dt \;+\; \frac{\beta}{2}\int_0^T \int_\Omega \rho\,|\boldsymbol{v}|^2\, dx\, dt \\ s.t. \quad & \partial_t\, \rho \;+\; \nabla_\Omega \cdot (\rho\, \boldsymbol{v}) \;=\; 0 \qquad \textit{in } Q := \Omega \times [0,T] \\ & \rho \geq 0 \; . \end{aligned}$$

with $\alpha, \beta \geq 0$. This model is a 4D model for joint image reconstruction, space-time total variation regularization and optimal transport. With the same arguments as in the previous section the momentum formulation of this model is a convex problem, since the total variation is convex. Due to the space-time TV regularization term we have an a priori estimate for

$$\rho \ \text{ in } L^p(0,T;BV(\Omega)) \ .$$

From an analytical as well as from the application point of view, the choice of p in the outer L^p norm is an interesting question. Can we expect a solution for $p = 1$? Now we proceed with verifying the existence of a solution of the problem aiming this question.

10.3.1 Existence

Theorem 10.3.2 (Existence of a Minimizer)**.** *Let $p > 0$. Then there exists a density function $\rho \in L^p(0,T;BV(\Omega))$ and a velocity field $\boldsymbol{v}$ as a solution of the constraint optimization problem in Model 10.3.1.*

Proof. To prove the existence of a solution $\rho, \mathbf{v}$ of our 4D-TV model above we have to prove strong compactness of sublevel-sets. For an arbitrary level $c \in \mathbb{R}$ we consider density functions ρ from the sublevel-set

$$S_c := \left\{ \rho \ \mid \int_0^T (\, |\rho|_{BV(\Omega)} \,)^p \ dt \leq c \right\}$$

in the following. A key idea of this proof is the simple transformation $\mathbf{w} := \sqrt{\rho}\, \mathbf{v}$. Hence, the third term in the functional above, coupling ρ and $\mathbf{v}$, can be rewritten as

$$\int_0^T \int_\Omega \rho \, |\mathbf{v}|^2 \ = \ \int_0^T \int_\Omega |\mathbf{w}|^2 \ .$$

This means we have an a-priori estimate for

$$w \ \text{ in } L^2((0,T) \times \Omega) \ . \tag{10.7}$$

In other words, we obtain a weak sequential compactness for w. The space-time regularization term, which depends the total variation in space, implies an a-priori estimate for

$$\rho \ \text{ in } L^p(0,T;BV(\Omega)) \ .$$

In addition, for the time derivative $\partial_t \rho$ an a-priori estimate is given by

$$\partial_t \rho \ \text{ in } L^q(0,T;W^{-1,s}(\Omega)) \ .$$

Then for $r > 1$ with $r < \frac{d}{d-1}$ we obtain $BV \subset L^r \subset W^{-1,s}$ with the compact embedding $BV \hookrightarrow L^r$ and the continuous embedding $L^r \hookrightarrow W^{-1,s}$. An adaption of the *Aubin-Lions Lemma* in [169, Chapter III, Prop. 1.3], gives **strong compactness** of

$$\rho \ \text{ in } L^p(0,T;L^r(\Omega)) \ . \tag{10.8}$$

The Lemma and its proof can be extended to the case of BV embeddings and weak* convergence instead of weak convergence. Thus (10.8) yields the **strong compactness** of the square-root of ρ,

$$\sqrt{\rho} \ \text{ in } L^{2p}(0,T;L^{2r}(\Omega)) \ . \tag{10.9}$$

Finally we can combine (10.7) and (10.9) and get a weak convergence of the product

$$\sqrt{\rho}\, w \ \text{ in } L^q(0,T;L^s(\Omega)) \ .$$

In order to prove the compactness of sublevel-sets we have to find an estimate for the latter coupling term $\sqrt{\rho}w$ in its corresponding space-time norm. In other words, keeping the a-priori estimates from above and an sublevel element ρ in mind, we need a condition for q, such that

$$\int_0^T \left(\int_\Omega |\, \sqrt{\rho}\, \mathbf{w}\, |^s \, dx \right)^{\frac{q}{s}} dt \ = \ \int_0^T \left(\int_\Omega |\, |\mathbf{w}|^s \cdot \rho^{\frac{s}{2}} \,| \, dx \right)^{\frac{q}{s}} dt$$

can be bounded from above. Obviously, $\frac{2}{s}$ and $\frac{2}{2-s}$ are Hölder conjugates,

$$\frac{1}{\left(\frac{2}{2-s}\right)} + \frac{1}{\left(\frac{2}{s}\right)} = 1 \ .$$

Hence, Hölder´s inequality applied to the inner integral yields

$$\begin{aligned}
&\int_0^T \left(\int_\Omega |\, |\mathbf{w}|^s \cdot \rho^{\frac{s}{2}} \,| \, dx \right)^{\frac{q}{s}} dt \\
\overset{\text{Hölder}}{\leq} &\int_0^T \left(\left(\int_\Omega (|\, \mathbf{w}\, |^s)^{\frac{2}{s}} \, dx \right)^{\frac{s}{2}} \cdot \left(\int_\Omega \left(\rho^{\frac{s}{2}}\right)^{\frac{2}{2-s}} \, dx \right)^{\frac{2-s}{2}} \right)^{\frac{q}{s}} dt \\
= &\int_0^T \left(\int_\Omega |\, \mathbf{w}\, |^2 \, dx \right)^{\frac{q}{2}} \cdot \left(\int_\Omega \rho^{\frac{s}{2-s}} \, dx \right)^{\frac{(2-s)q}{2s}} dt \ .
\end{aligned}$$

Applying Hölder´s inequality with the conjugates $\frac{2}{q}$ and $\frac{2}{2-q}$ with

$$\frac{1}{\left(\frac{2}{q}\right)} + \frac{1}{\left(\frac{2}{2-q}\right)} = 1$$

once more, we obtain

$$\overset{\text{Hölder}}{\leq} \underbrace{\left(\int_0^T \int_\Omega |\,\mathbf{w}\,|^2 \, dx\, dt\right)^{\frac{q}{2}}}_{<\ \infty} \cdot \left(\int_0^T \left(\int_\Omega \rho^{\frac{s}{2-s}}\, dx\right)^{\frac{(2-s)q}{s(2-q)}} dt\right)^{\frac{2-q}{2}} .$$

The left factor is bounded since $\mathbf{w} \in L^2((0,T)\times\Omega)$, see 10.7. Now we concentrate on the space-time integral in the right factor. We know ρ is in $L^p(0,T;L^r(\Omega))$, see 10.8. That is why we need $\frac{s}{2-s} \overset{!}{=} r$ and $\frac{(2-s)q}{s(2-q)} \overset{!}{=} p$. Combining these conditions we can deduce conditions on s, q and p to get strong compactness:

$$r \overset{!}{=} \frac{s}{2-s} \;\Rightarrow\; s = \frac{2r}{1+r} > 1, \text{ since } r > 1 .$$

$$p \overset{!}{=} \frac{(2-s)q}{s(2-q)} = \frac{q}{(2-q)r} \;\Rightarrow\; p\,r\,(2-q) = q$$
$$\Rightarrow\; q = \frac{2\,p\,r}{1+p\,r} > 1$$

This means we obtain $q > 1$ if $p\,r > 1$, which in turn is fulfilled if $p > 1$ since $r > 1$. Finally we can conclude strong compactness for $\sqrt{\rho}w$ in $L^q(0,T;L^s(\Omega))$ and therewith the existence of a solution if p is chosen larger than one. □

Interpretation:
Hence we could see that the analysis of existence relies on the superlinear growth of the TV regularization term inside the time integral

$$\int_0^T (\ |\rho|_{BV(\Omega)}\ )^p \ dt \ ,$$

although $p = 1$ seems a reasonable choice. Choosing $p = 1$ would imply that we seek a minimal L^1 norm of the vector of total variations. Similarly to the standard spatial case of TV this strategy favors sparsity, i.e. solutions with very large total variation are allowed at some time step, whereas small total variation is favored in the other time steps. In the extreme case, one can think of staircasing in space and time. For some applications time dependent sparsity regularization could be very useful. For example in meteorology, one can be interested in analyzing sharp shock fronts for weather forecast due to time, see for instance [73]. In the following sections on numerical results of this model we will give some indications of this effect.

10.3.2 Uniqueness

Now we pass over to the uniqueness of a solution $\rho, \mathbf{v}$ of Model 10.3.1. First of all, we reformulate our problem in Model (10.3.1) via the momentum substitution $\mathbf{j} := \rho\mathbf{v}$, and obtain an equivalent problem dependent on ρ and $\mathbf{j}$.

Model 10.3.3. *Joint 4D Model with Space-Time TV Regularization*

$$\begin{aligned} \min_{\rho,\boldsymbol{j}} \; & \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Sigma)}\,dt \;+\; \frac{\alpha}{p}\int_0^T (\,|\rho|_{BV(\Omega)}\,)^p\,dt \;+\; \frac{\beta}{2}\int_0^T\int_\Omega \frac{|\,\boldsymbol{j}\,|^2}{\rho}\,dx\,dt \\ s.t. \quad & \partial_t\,\rho \;+\; \nabla_\Omega\cdot\boldsymbol{j} \;=\; 0 \qquad \text{in } Q := \Omega\times[0,T] \\ & \rho \geq 0\;. \end{aligned}$$

This optimization problem is strictly convex, because the functional is strictly convex and the constraint is linear, hence convex. For the proof we basically follow concepts of uniqueness proofs in mean-field games, see [114].

Theorem 10.3.4 (Uniqueness of a Minimizer)**.** *Let $p > 0$. Then there exists at most one density function $\rho \in L^p(0,T;BV(\Omega))$ and at most one velocity field $\boldsymbol{v}$ as a minimizer of the constraint optimization problem in Model 10.3.1.*

Proof. The Lagrangian of the new constrained optimization problem is given by

$$L(\rho,\mathbf{j};\lambda) \;=\; \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Omega)}\,dt \;+\; \frac{\alpha}{p}\int_0^T (\,|\rho|_{BV(\Omega)}\,)^p\,dt \tag{10.10}$$

$$+\; \frac{\beta}{2}\int_0^T\int_\Omega \frac{|\,\mathbf{j}\,|^2}{\rho}\,dx\,dt \;+\; \langle\,\lambda\,,\,\partial_t\,\rho \;+\; \nabla_\Omega\cdot\mathbf{j}\,\rangle_Q\;, \tag{10.11}$$

where λ denotes a Lagrange multiplier function for the mass conservation constraint. Note that the existence of a Lagrange multiplier function is not obvious.
However, if we assume the existence of a Lagrange multiplier, we can formally write the optimality system for an optimal solution $(\rho,\mathbf{j},\lambda)$, i.e. the Euler-Lagrange equations for the Lagrangian above, are given by

$$\frac{\partial L}{\partial \mathbf{j}} \;=\; \beta\,\frac{\mathbf{j}}{\rho} - \nabla\lambda \;=\; 0 \tag{10.12}$$

$$\frac{\partial L}{\partial \rho} \;=\; \tilde{K}^*\left(\tilde{K}\rho - f\right) \;+\; \xi \;-\; \frac{\beta\,|\,\mathbf{j}\,|^2}{2\,\rho^2} \;-\; \partial_t\,\lambda \;=\; 0 \tag{10.13}$$

with a subgradient $\xi \in \partial L_2(\rho)$, where L_2 is the TV regularization term in the Lagrangian including α. We assume a rescaling of the Lagrange multiplier λ,

$$\lambda(0) = \lambda(T) = 0\;.$$

The first equation of our optimality system (10.12) is equivalent to

$$\mathbf{j} = \frac{1}{\beta}\,\rho\,\nabla\lambda\ . \tag{10.14}$$

Thus we can plug (10.12) in (10.13) and obtain a combined form of the optimality system:

$$\underbrace{\tilde{K}^*\tilde{K}\rho - \tilde{K}^* f\ +\ \xi}_{=:\,q\,\in\,\partial(L_1+L_2)(\rho)} - \frac{1}{2}|\nabla\lambda|^2\ -\ \partial_t\,\lambda\ =\ 0\ , \tag{10.15}$$

where L_1 denotes the data fidelity in the Lagrangian above. The equation (10.15) does no longer depend on $\mathbf{j}$ directly, but an optimal $\mathbf{j}$ fulfills the mass conservation constraint in our original problem (10.3.1). Using (10.14) we obtain another optimality equation regarding ρ and the dual variable λ,

$$\partial_t\,\rho\ +\ \nabla_\Omega\cdot(\rho\,\nabla\lambda)\ =\ 0\ . \tag{10.16}$$

With (10.15) and (10.15) we now have two equations dependent on ρ and λ only. Let $(\rho_1, \mathbf{j}_1, \lambda_1)$ and $(\rho_2, \mathbf{j}_2, \lambda_2)$ be two solutions of our minimization problem in (10.11). To prove uniqueness we show that these solutions are one and the same. Writing the equations in (10.15) and (10.16) for both solutions and subtracting the corresponding ones we get

$$\begin{aligned} q_1 - q_2\ -\ \frac{1}{2}\left(|\nabla\lambda_1|^2 - |\nabla\lambda_2|^2\right)\ -\ \partial_t\left(\lambda_1-\lambda_2\right)\ &=\ 0\\ \partial_t\left(\rho_1-\rho_2\right)\ +\ \nabla_\Omega\cdot\left(\rho_1\,\nabla\lambda_1 - \rho_2\,\nabla\lambda_2\right)\ &=\ 0 \end{aligned}$$

Multiplying the first equation with $(\rho_1 - \rho_2)$, multiplying the second equation with $(\lambda_1 - \lambda_2)$, integrating both equations over $\Omega\times(0,1)$ and subtracting them we can deduce

$$\begin{aligned} 0\ \ =\ \ &\langle q_1 - q_2\ ,\ \rho_1-\rho_2\rangle_Q\ -\ \frac{1}{2}\left\langle |\nabla\lambda_1|^2 - |\nabla\lambda_2|^2\ ,\ \rho_1-\rho_2\right\rangle_Q && (10.17)\\ -\ &\langle\partial_t\left(\lambda_1-\lambda_2\right)\ ,\ \rho_1-\rho_2\rangle_Q\ -\ \langle\partial_t\left(\rho_1-\rho_2\right)\ ,\ \lambda_1-\lambda_2\rangle_Q && (10.18)\\ +\ &\langle\rho_1\,\nabla\lambda_1 - \rho_2\,\nabla\lambda_2\ ,\ \nabla\lambda_1 - \nabla\lambda_2\rangle_Q\ . && (10.19) \end{aligned}$$

The last addend results from using Gauss's theorem. Applying Gauss's theorem in line (10.18) once more, the terms cancel each other. The first term in (10.17) is a symmetric Bregman distance regarding the functional with the data fidelity and the TV regularization term $L_1 + L_2$, i.e.

$$\langle q_1 - q_2\ ,\ \rho_1-\rho_2\rangle_Q\ =\ D^{symm}_{L_1+L_2}(\rho_1,\rho_2)\ \geq\ 0\ .$$

Hence the resulting equality reads as follows

$$\begin{aligned} 0\ =\ D^{symm}_{L_1+L_2}(\rho_1,\rho_2) + \frac{1}{2}\int_Q\Big(\ &\rho_2\left(|\nabla\lambda_1|^2 - |\nabla\lambda_2|^2\right) - \rho_1\left(|\nabla\lambda_1|^2 - |\nabla\lambda_2|^2\right)\\ &+2\rho_1\left(|\nabla\lambda_1|^2 - \nabla\lambda_1\nabla\lambda_2\right) - 2\rho_2\left(\nabla\lambda_1\nabla\lambda_2 - |\nabla\lambda_2|^2\right)\ \Big)\ . \end{aligned}$$

Application of the binomial theorem implies

$$0 \;=\; D^{symm}_{L_1+L_2}(\rho_1,\rho_2) \;+\; \frac{1}{2}\int_Q \underbrace{(\rho_1+\rho_2)\,(\nabla\lambda_1-\nabla\lambda_2)^2}_{\geq\,0}\;.$$

To identify our two solutions ρ_1 and ρ_2 with each other, we use some error estimation techniques and the linearity of K:

$$\begin{aligned}
&\geq\; D^{symm}_{L_1+L_2}(\rho_1,\rho_2)\\
&=\; \langle q_1-q_2\;,\;\rho_1-\rho_2\rangle_Q\\
&\overset{(10.15)}{=}\; \left\langle \tilde{K}^*\tilde{K}\,(\rho_1-\rho_2)\;,\;\rho_1-\rho_2\right\rangle_Q \;+\; \left\langle\left(-\tilde{K}^*f_1+\xi_1\right)-\left(-\tilde{K}^*f_2+\xi_2\right)\;,\;\rho_1-\rho_2\right\rangle_Q\\
&=\; \underbrace{\left\|\tilde{K}\,(\rho_1-\rho_2)\right\|_2^2}_{\geq\,0} \;+\; \underbrace{D^{symm}_{\left(-\langle\tilde{K}\,\cdot,f\rangle+L_2\right)}(\rho_1,\rho_2)}_{\geq\,0}\;.
\end{aligned}$$

If the linear compact operator K is injective, we can finally conclude $\rho_1=\rho_2$ and $\mathbf{j}_1=\mathbf{j}_2$ using (10.14) - (10.16). □

10.4 Numerical Realization: Newton-SQP for the L^2 case

In this section we present a Newton-SQP algorithm for solving Model 10.2.1. From a numerical point of view it is reasonable to reformulate the model via the momentum substitution $\mathbf{j}:=\rho\mathbf{v}$. With this substitution the system reads as follows

Model 10.4.1. *Joint 4D Model with L^2 Regularization*

$$\begin{aligned}
&\min_{\rho,\,\boldsymbol{j}}\; \frac{1}{2}\int_0^T \|K\rho-f\|^2_{L^2(\Sigma)}\,dt \;+\; \frac{\alpha}{2}\int_0^T \|\nabla\rho\|_2^2\;dt \;+\; \frac{\beta}{2}\int_0^T\int_\Omega \frac{|\,\boldsymbol{j}\,|^2}{\rho}\,dx\,dt\\
&s.t.\quad c(\boldsymbol{j},\rho) \;:=\; \nabla_Q\cdot\begin{pmatrix}\boldsymbol{j}\\ \rho\end{pmatrix} \;=\; 0 \qquad in\ Q:=\Omega\times[0,T]\\
&\qquad\quad \rho\geq 0\;.
\end{aligned}$$

This problem is convex. The objective functional is convex, since the coupling term is a quadratic-over-linear function, see 3.1.1, and the constraints are convex since the mass conservation constraint is linear in ρ, respectively $\mathbf{j}$. The continuity equation "reduces" to a space-time divergence for the joint function $(\mathbf{j},\rho)$ which is very useful for concise implementations. Newton-SQP methods are based on the Jacobian of

the optimality system. Hence, they can be quadratically convergent and are very useful for computing solutions of high accuracy.

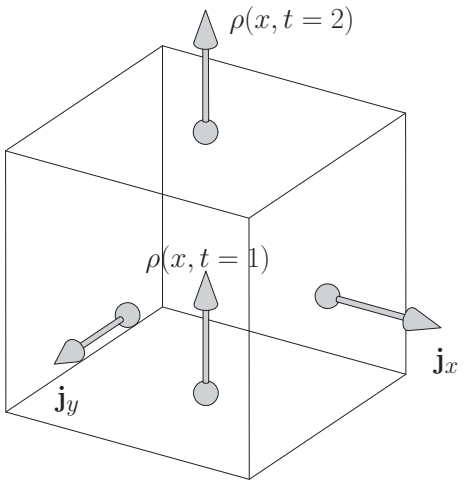
In this section we follow the concept "*first discretize and then optimize*", i.e. we first discretize the objective functional and the constraints and then build the discretized Karush-Kuhn-Tucker (KKT) optimality system which is solved numerically.

10.4.1 Optimality Conditions - KKT System

Discretization:
In following let us assume $\Omega \subset \mathbb{R}^2$ for the spatial domain to simplify the notation and illustration of the discretization. In the following we assume the step size

$$h_x = \frac{1}{n_x}, h_y = \frac{1}{n_y} \text{ and } h_t = \frac{1}{n_t} ,$$

where n_x and n_y denote the spacial dimension and n_t denotes the temporal dimension. Let $h := h_x\, h_y\, h_t$. In order to derive a stable discretization of our objective functional and the constraint, we define the density ρ and the momentum $\mathbf{j}$ on a staggered grid, see Figure 10.4.1.



Hence, in order to evaluate ρ and $\mathbf{j}$ in the cell center, we need to define averaging operators. With the averaging stencil,

$$A_{n_x} = \frac{1}{2} \begin{pmatrix} 1 & 1 & & & \\ & 1 & 1 & & 0 \\ & & \ddots & \ddots & \\ & 0 & & 1 & 1 \\ & & & & 1 & 1 \end{pmatrix} \in \mathbb{R}^{n_x \times (n_x - 1)} ,$$

the spacial averaging matrix for the momentum is defined as

$$A_\Omega := [A_x \; A_y] \ ,$$

where

$$A_x = I_{n_t} \otimes \left(I_{n_y} \otimes A_{n_x}\right) \qquad \text{and} \qquad A_y = I_{n_t} \otimes \left(A_{n_y} \otimes I_{n_x}\right) \ .$$

The temporal averaging matrix for the density ρ is defined as

$$A_t = A_{n_t} \otimes \left(I_{n_y} \otimes I_{n_x}\right) \ .$$

In order to define derivatives in space, we take the forward difference matrix with Neumann boundary conditions

$$D_{n_x} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & 0 \\ & & \ddots & \ddots & \\ & 0 & & -1 & 1 \\ & & & & -1 \quad 1 \end{pmatrix} \in \mathbb{R}^{n_x \times (n_x - 1)} \ ,$$

and thus can define the discretized gradient matrix in space as

$$\mathbf{D_\Omega} := \begin{pmatrix} I_{n_y} \otimes D_{n_x} \\ D_{n_y} \otimes I_{n_x} \end{pmatrix} \ .$$

Then $(-\mathbf{D_\Omega})^T \mathbf{D_\Omega}$ is the central difference matrix for second order derivatives in space. Now, we have the main ingredients to define the discretized objective functional and the discretized constraint.

For $\alpha, \beta \geq 0$ the discretized version of the objective functional in Model 10.4.1 reads as follows

$$F(\mathbf{j}, \rho) := h \left[\sum_t \sum_{i,j} \left[R.^2\right]_{ijt} + \frac{\alpha}{2} \tilde{\mathbf{D}}_{\mathbf{\Omega}}^T \tilde{\mathbf{D}}_{\mathbf{\Omega}} + \frac{\beta}{2} \sum_t \sum_{i,j} \left[A_\Omega(\mathbf{j}.^2) \odot A_t(\rho.^{-1})\right]_{ijt} \right] \ ,$$

where we used a time-extended spatial residual vector R and a time-extended spatial derivative $\tilde{\mathbf{D}}_{\mathbf{\Omega}}$, defined as

$$R = \begin{pmatrix} [K\rho - f]_{t=1} \\ \vdots \\ [K\rho - f]_{t=n_t+1} \end{pmatrix} \qquad \text{and} \qquad \tilde{\mathbf{D}}_{\mathbf{\Omega}} = \begin{pmatrix} \mathbf{D_\Omega} [\rho]_{t=1} \\ \vdots \\ \mathbf{D_\Omega} [\rho]_{t=n_t+1} \end{pmatrix} \ .$$

In order to discretize the constraint (the continuity equation) in Model 10.4.1, we define the divergence matrix in space-time, i.e.

$$\mathbf{D_Q} := h \left[\mathbf{D_x}, \mathbf{D_y}, \mathbf{D_t}\right] \ ,$$

where

$$\mathbf{D_x} = I_{n_t} \otimes \left(I_{n_y} \otimes \frac{1}{h_x} D_{n_x+1} \right) ,$$

$$\mathbf{D_y} = I_{n_t} \otimes \left(\frac{1}{h_y} D_{n_y+1} \otimes I_{n_x} \right) \qquad \text{and} \qquad \mathbf{D_t} = \frac{1}{h_t} D_{n_t+1} \otimes \left(I_{n_y} \otimes I_{n_x} \right) ,$$

denote the discretized derivatives in space-time. Hence, the discretized constraint simply is given by

$$C(\mathbf{j}, \rho) \; = \; \mathbf{D_Q} \begin{pmatrix} \mathbf{j} \\ \rho \end{pmatrix} .$$

Thus, we can formulate the Lagrangian of the system,

$$L(\mathbf{j}, \rho, \lambda) \; = \; F(\mathbf{j}, \rho) \; + \; (C(\mathbf{j}, \rho))^T \, \lambda \, ,$$

where $\lambda \geq 0$ denotes a Lagrange multiplier due to the constraint. With

$$dR \; = \; \begin{pmatrix} \left[K^T \left(K\rho - f \right) \right]_{t=1} \\ \vdots \\ \left[K^T \left(K\rho - f \right) \right]_{t=n_t+1} \end{pmatrix} \qquad \text{and} \qquad \tilde{\Delta} \; = \; \begin{pmatrix} \mathbf{D_\Omega}^T \mathbf{D_\Omega} \left[\rho \right]_{t=1} \\ \vdots \\ \mathbf{D_\Omega}^T \mathbf{D_\Omega} \left[\rho \right]_{t=n_t+1} \end{pmatrix} ,$$

an optimal solution $z := (\mathbf{j}, \rho, \lambda)$ of our problem minimizes the Lagrangian and fulfills the following

KKT system:

$$0 \; = \; \begin{pmatrix} L_{\mathbf{j}}(\mathbf{j}, \rho, \lambda) \\ L_{\rho}(\mathbf{j}, \rho, \lambda) \\ L_{\lambda}(\mathbf{j}, \rho, \lambda) \end{pmatrix} = \begin{pmatrix} h \left[\, \beta \, \mathrm{diag}(\mathbf{j}) \, A_\Omega^T A_t \left(\rho.^{-1} \right) \, \right] \; + \; (C(\mathbf{j}, \rho))^T \, \lambda \\ h \left[\, dR \; - \; \alpha \tilde{\Delta} \; - \; \beta \, \mathrm{diag}(\rho.^{-2}) \, A_\Omega^T A_t \left(\mathbf{j}.^2 \right) \, \right] \\ C(\mathbf{j}, \rho) \end{pmatrix} := F(z) \, .$$

The positivity constraint of our density ρ can be controlled via a selection rule in a line search strategy for the Newton step size.

10.4.2 Newton-SQP Algorithm

With $z = (\mathbf{j}, \rho, \lambda)$ a Newton iteration for determining zeros of F, resulting from Taylor linearization, reads as follows

$$z^{k+1} = z^k - F'(z^k)^{-1} \cdot F(z^k) \, .$$

Equivalently, we have to solve the following system of linear equations in each iteration step

$$\begin{pmatrix} L_{\mathbf{jj}}(\mathbf{j}^k, \rho^k, \lambda^k) & L_{\mathbf{j}\rho}(\mathbf{j}^k, \rho^k, \lambda^k) & C'(\mathbf{j}^k, \rho^k)^T \\ L_{\rho\mathbf{j}}(\mathbf{j}^k, \rho^k, \lambda^k) & L_{\rho\rho}(\mathbf{j}^k, \rho^k, \lambda^k) & \\ C'(\mathbf{j}^k, \rho^k) & & 0 \end{pmatrix} \begin{pmatrix} d_{\mathbf{j}}^k \\ d_\rho^k \\ d_\lambda^k \end{pmatrix} = - \begin{pmatrix} L_{\mathbf{j}}(\mathbf{j}^k, \rho^k, \lambda^k) \\ L_{\rho}(\mathbf{j}^k, \rho^k, \lambda^k) \\ L_{\lambda}(\mathbf{j}^k, \rho^k, \lambda^k) \end{pmatrix} , \tag{10.20}$$

where $d_{\mathbf{j}}$ denotes the search direction regarding $\mathbf{j}$.
For solving this system of linear equations in each Newton iteration, we use a preconditioned GMRES solver, implemented in MATLAB. After computing the new search directions $d_z^k = (d_{\mathbf{j}}^k, d_\rho^k, d_\lambda^k)$ we obtain the new primal and dual iterates by the following updates

$$\begin{aligned} \mathbf{j}^{k+1} &= \mathbf{j}^k + \sigma^k d_{\mathbf{j}}^k \\ \rho^{k+1} &= \rho^k + \sigma^k d_\rho^k \\ \lambda^{k+1} &= \lambda^k + \sigma^k d_\lambda^k \,. \end{aligned}$$

This method is called Newton-SQP method because we have to solve a sequence of quadratic problems. In the k-th Newton step the system of equations above (10.20) represents the optimality condition of a quadratic problem with linear constraints.

10.4.3 Line-Search and Multigrid Preconditioning

Positivity via Line-Search
The iteration point z^k is not necessarily feasible for the original problem for example regarding positivity and d^k is in general not a feasible descent direction of the objective in z^k. Therefore the step size σ in each step is determined via a line search strategy. We use a merit function and an Armijo type rule to obtain an efficient step size. Thus we get a sufficient descent of the merit function in z^k. The initial step size for the line search strategy is chosen such that positivity of ρ is guaranteed.

Preconditioning via Multigrid
For the solution of the large linear system of equations in each Newton step, we used the GMRES solver implemented in MATLAB with different preconditioning strategies. We compared a simple Schur complement preconditioner, a preconditioner presented in [85] and a multigrid preconditioner. We implemented the multigrid with V-cycles and a relaxation via SOR. For our test cases the multigrid preconditioner worked best.

The Newton-SQP method can be quadratically convergent. Compared to the first order splitting methods in the following section we only needed 20 preconditioned Newton iteration steps for most of our test cases to reach a sufficient accuracy. However, for higher dimensional 3D and 4D data sets the sub-steps in the preconditioned GMRES solver reached memory limits quite early and the cost per iteration was quite high. This could be improved by the application of domain decomposition methods or the parallelization of GMRES on GPUs.

10.5 Numerical Realization: Splitting Methods for the TV case

In this section we will present an efficient splitting method to solve the joint 4D reconstruction Model 10.3.1 with spatio-temporal TV regularization. Our goal is to find a numerical splitting algorithm, where each of the resulting sub-steps can be computed very efficiently, needs only a small amount of system resources and may be parallelized. To derive and realize this numerical strategy, we make extensive use of the concepts we presented in Chapter 4 and computational science.

For computational reasons we follow the idea of split Bregman and further substitute the gradient in the TV terms for an auxiliary function z. Hence, an equivalent constrained optimization problem that is equivalent to Model 10.3.1 reads as follows

$$\begin{aligned}
&\min_{\rho,\mathbf{j}} \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Omega)}\,dt \;+\; \frac{\alpha}{p}\int_0^T \Big(\int_\Omega |z|\;dx\Big)^p\;dt \;+\; \frac{\beta}{2}\int_0^T\int_\Omega \frac{|\nu|^2}{w}\,dx\,dt \\
&\text{s.t.}\quad \partial_t\,\rho \;+\; \nabla_\Omega\cdot\mathbf{j} \;=\; 0 \qquad \text{in } Q \\
&\qquad\quad \forall t\in[0,T]:\;\; \nabla_\Omega\,\rho = z \qquad \text{in } \Omega \\
&\qquad\quad w = \rho \qquad \text{in } Q \\
&\qquad\quad \nu = \mathbf{j} \qquad \text{in } Q \\
&\qquad\quad \rho \geq 0 \qquad \text{in } Q\;.
\end{aligned}$$

It turned out that the coupling of the momentum $\mathbf{j}$ and the density ρ in the third term of the objective functional is a difficulty concerning the derivation of efficient splitting schemes. Thus, in order to obtain a sufficiently decoupled system for the momentum $\mathbf{j}$ and the density function ρ, we decided to introduce additional auxiliary functions w and ν.

By introducing dual variables, Lagrange multiplier, the Lagrangian for this decoupled system reads as follows:

$$\begin{aligned}
&L(\rho,\mathbf{j},z,w,\nu;\;q,y,r,\eta) \\
&= \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Omega)}\,dt \;+\; \frac{\alpha}{p}\int_0^T \Big(\int_\Omega |z|\;dx\Big)^p\;dt \;+\; \frac{\beta}{2}\int_0^T\int_\Omega \frac{\nu^2}{w}\,dx\,dt \\
&\quad + \langle q\,,\,\partial_t\,\rho \;+\; \nabla_\Omega\cdot\mathbf{j}\rangle_Q \;+\; \int_0^T \langle y\,,\,\nabla_\Omega\,\rho - z\rangle_\Omega\;dt \;+\; \langle r\,,\,w-\rho\rangle_Q \;+\; \langle \eta\,,\,\nu-\mathbf{j}\rangle_Q
\end{aligned}$$

where $\rho,\mathbf{j},z,w$ and ν denote the primal variables, q,y,r and η denote the Lagrange multipliers respectively the dual variables. Following the idea of augmented Lagrangian methods and inexact Uzawa algorithms we define the corresponding augmented Lagrangian L^μ with an additional preconditioning norm for the primal function ρ.

Augmented Lagrangian:

$$
\begin{aligned}
&L^{\mu}(\rho, \rho^k, \mathbf{j}, z, w, \nu;\ q, y, r, \eta) \\
&= \frac{1}{2}\int_0^T \|K\rho - f\|^2_{L^2(\Omega)}\, dt \ + \ \frac{\alpha}{p}\int_0^T \Big(\int_\Omega |z| \ dx \Big)^p \ dt \ + \ \frac{\beta}{2}\int_0^T \int_\Omega \frac{\nu^2}{w}\, dx\, dt \\
&\quad + \ \langle q \ , \ \partial_t \rho \ + \ \nabla_\Omega \cdot \mathbf{j} \rangle_Q \ + \ \frac{\mu_1}{2}\, \|\partial_t \rho \ + \ \nabla_\Omega \cdot \mathbf{j}\|^2_{L^2(Q)} \\
&\quad + \ \int_0^T \Big(\langle y \ , \ \nabla_\Omega \rho - z \rangle_\Omega \ + \ \frac{\mu_2}{2}\, \|z - \nabla_\Omega \rho\|^2_{L^2(Q)} \Big)\, dt \\
&\quad + \ \langle r \ , \ w - \rho \rangle_Q \ + \ \frac{\mu_3}{2}\, \|w - \rho\|^2_{L^2(Q)} \\
&\quad + \ \langle \eta \ , \ \nu - \mathbf{j} \rangle_Q \ + \ \frac{\mu_4}{2}\, \|\nu - \mathbf{j}\|^2_{L^2(Q)} \ + \ \frac{1}{2}\, \big\|\rho - \rho^k\big\|^2_{P_\delta(Q)} \ .
\end{aligned}
$$

Similarly to the alternative direction minimization algorithm (ADMM), Douglas-Rachford splitting or Split Bregman (see Subsection 4.4.2), the basic idea of the following splitting strategies is to alternatively minimize the augmented Lagrangian with respect to the primal variables and to update the Lagrange multipliers subsequently. The main difference of our approach compared to these methods is the introduction of preconditioning techniques in the sense of inexact Uzawa algorithms (or Bregmanized operator splitting (BOS), see Subsection 4.4.3).

10.5.1 Inexact Uzawa & Bregmanized Operator Splitting

The basic idea of inexact Uzawa methods respectively Bregmanized operator splitting is to add a preconditioning term to the augmented Lagrangian, see the last term in the augmented Lagrangian above, which depends on P_δ,

$$
\big\|\rho - \rho^{k+1}\big\|^2_{P_\delta} = \big\langle \rho - \rho^k, P_\delta(\rho - \rho^k) \big\rangle \ .
$$

This term relates subsequent primal iterates in a specific, controllable way.

Hence, the following splitting algorithm is an inexact Uzawa method applied to the augmented Lagrangian:

Algorithm 10.5.1 (Inexact Uzawa Algorithm for Model 10.3.1)**.**

Primal updates:

$$\rho^{k+1} = \arg\min_{\rho} \; L^{\mu}(\rho, \rho^k, \boldsymbol{j}^k, z^k, w^k, \nu^k, q^k, y^k, r^k, \eta^k) \tag{I}$$

$$\boldsymbol{j}^{k+1} = \arg\min_{\boldsymbol{j}} \; L^{\mu}(\rho^{k+1}, \rho^k, \boldsymbol{j}, z^k, w^k, \nu^k, q^k, y^k, r^k, \eta^k) \tag{II}$$

$$z^{k+1} = \arg\min_{z} \; L^{\mu}(\rho^{k+1}, \rho^k, \boldsymbol{j}^{k+1}, z, w^k, \nu^k, q^k, y^k, r^k, \eta^k) \tag{III}$$

$$w^{k+1} = \arg\min_{w} \; L^{\mu}(\rho^{k+1}, \rho^k, \boldsymbol{j}^{k+1}, z^{k+1}, w, \nu^k, q^k, y^k, r^k, \eta^k) \tag{IV}$$

$$\nu^{k+1} = \arg\min_{\nu} \; L^{\mu}(\rho^{k+1}, \rho^k, \boldsymbol{j}^{k+1}, z^{k+1}, w^{k+1}, \nu, q^k, y^k, r^k, \eta^k) \tag{V}$$

Dual updates:

$$q^{k+1} = q^k \; + \; \mu_1 \left(\partial_t \rho^{k+1} + \nabla_\Omega \cdot \boldsymbol{j}^{k+1}\right) \tag{VI}$$

$$y^{k+1} = y^k \; + \; \mu_2 \left(\nabla_\Omega \, \rho^{k+1} - z^{k+1}\right) \qquad \forall t \in [0,T] \tag{VII}$$

$$r^{k+1} = r^k \; + \; \mu_3 \left(w^{k+1} - \rho^{k+1}\right) \tag{VIII}$$

$$\eta^{k+1} = \eta^k \; + \; \mu_4 \left(\nu^{k+1} - \boldsymbol{j}^{k+1}\right) \tag{IX}$$

The dual update can be computed efficiently, because they only depend on simple differential operators that are applied. For the rest of this subsection we will concentrate on the minimization problems in the primal updates.

By transferring terms from the scalar products to the L^2 norms and by leaving terms that are independent of the minimization function, we obtain the following variational problems, which need to be solved in each iteration,

$$\begin{aligned}
\rho^{k+1} = \arg\min_{\rho} \Bigg\{ & \frac{1}{2} \int_0^T \|K\rho - f\|^2_{L^2(\Omega)} \, dt \; + \; \left\langle q^k, \partial_t \rho \right\rangle_Q \; + \; \frac{\mu_1}{2} \left\| \partial_t \rho + \nabla_\Omega \cdot \mathbf{j}^k \right\|^2_{L^2(Q)} \\
& + \int_0^T \left(\left\langle y^k \, , \, \nabla_\Omega \, \rho \right\rangle_\Omega \; + \; \frac{\mu_2}{2} \left\| \nabla_\Omega \, \rho - z^k \right\|^2_{L^2(\Omega)} \right) dt \; + \; \left\langle r^k \, , \, -\rho \right\rangle_Q \\
& + \frac{\mu_3}{2} \left\| \rho - w^k \right\|^2_{L^2(Q)} \; + \; \frac{1}{2} \left\| \rho - \rho^k \right\|^2_{P_\delta(Q)} \Bigg\}
\end{aligned} \tag{I}$$

$$\begin{aligned}
= \arg\min_{\rho} \Bigg\{ & \frac{1}{2} \int_0^T \|K\rho - f\|^2_{L^2(\Omega)} \, dt \; + \; \frac{\mu_1}{2} \left\| \partial_t \rho + \nabla_\Omega \cdot \mathbf{j}^k + \frac{1}{\mu_1} q^k \right\|^2_{L^2(Q)} \\
& + \; \frac{\mu_2}{2} \int_0^T \left\| \nabla_\Omega \, \rho - z^k + \frac{1}{\mu_2} y^k \right\|^2_{L^2(\Omega)} dt \; + \; \frac{\mu_3}{2} \left\| \rho - w^k - \frac{1}{\mu_3} r^k \right\|^2_{L^2(Q)} \\
& + \frac{1}{2} \left\| \rho - \rho^k \right\|^2_{P_\delta(Q)} \Bigg\}
\end{aligned}$$

$$\begin{aligned}\mathbf{j}^{k+1} &= \arg\min_{\mathbf{j}} \Big\{ \big\langle q^k \,,\, \nabla_\Omega \cdot \mathbf{j} \big\rangle_Q \;+\; \frac{\mu_1}{2} \left\| \partial_t\, \rho^{k+1} \;+\; \nabla_\Omega \cdot \mathbf{j} \right\|^2_{L^2(Q)} \\ &\qquad + \big\langle \eta^k \,,\, -\mathbf{j} \big\rangle_Q \;+\; \frac{\mu_4}{2} \left\| \nu^k - \mathbf{j} \right\|^2_{L^2(Q)} \Big\} \\ &= \arg\min_{\mathbf{j}} \left\{ \frac{\mu_1}{2} \left\| \nabla_\Omega \cdot \mathbf{j} + \partial_t\, \rho^{k+1} + \frac{1}{\mu_1} q^k \right\|^2_{L^2(Q)} \;+\; \frac{\mu_4}{2} \left\| \mathbf{j} - \nu^k - \frac{1}{\mu_4} \eta^k \right\|^2_{L^2(Q)} \right\}\end{aligned} \tag{II}$$

$$\begin{aligned}z^{k+1} &= \arg\min_{z} \left\{ \int_0^T \left(\frac{\alpha}{p} \Big(\int_\Omega |z|\; dx \Big)^p \;+\; \big\langle y^k \,,\, -z \big\rangle_\Omega \;+\; \frac{\mu_2}{2} \left\| z - \nabla_\Omega\, \rho^{k+1} \right\|^2_{L^2(\Omega)} \right) dt \right\} \\ &= \arg\min_{z} \left\{ \int_0^T \left(\frac{\alpha}{p} \Big(\int_\Omega |z|\; dx \Big)^p \;+\; \frac{\mu_2}{2} \left\| z - \left(\nabla_\Omega\, \rho^{k+1} + \frac{1}{\mu_2} y^k \right) \right\|^2_{L^2(\Omega)} \right) dt \right\}\end{aligned} \tag{III}$$

$$\begin{aligned}w^{k+1} &= \arg\min_{w} \left\{ \frac{\beta}{2} \int_0^T \int_\Omega \frac{\left|\nu^k\right|^2}{w}\, dx\, dt \;+\; \big\langle r^k \,,\, w \big\rangle_Q \;+\; \frac{\mu_3}{2} \left\| w - \rho^{k+1} \right\|^2_{L^2(Q)} \right\} \\ &= \arg\min_{w} \left\{ \frac{\beta}{2} \int_0^T \int_\Omega \frac{\left|\nu^k\right|^2}{w}\, dx\, dt \;+\; \frac{\mu_3}{2} \left\| w - \rho^{k+1} + \frac{1}{\mu_3} r^k \right\|^2_{L^2(Q)} \right\}\end{aligned} \tag{IV}$$

$$\begin{aligned}\nu^{k+1} &= \arg\min_{z} \left\{ \frac{\beta}{2} \int_0^T \int_\Omega \frac{|\nu|^2}{w^{k+1}}\, dx\, dt \;+\; \langle \eta \,,\, \nu \rangle_Q \;+\; \frac{\mu_4}{2} \left\| \nu - \mathbf{j}^{k+1} \right\|^2_{L^2(Q)} \right\} \\ &= \arg\min_{z} \left\{ \frac{\beta}{2} \int_0^T \int_\Omega \frac{|\nu|^2}{w^{k+1}}\, dx\, dt \;+\; \frac{\mu_4}{2} \left\| \nu - \mathbf{j}^{k+1} + \frac{1}{\mu_4} \eta^k \right\|^2_{L^2(Q)} \right\} .\end{aligned} \tag{V}$$

The optimality conditions of these variational problems lead to the following update formulas of the primal variables.

Primal updates:

$$\begin{aligned}\rho^{k+1} &= \left(\tilde{K}^* \tilde{K} + \mu_3 - \mu_1 \partial_t^2 - \mu_2 \Delta_\Omega + P_\delta \right)^{-1} \\ &\qquad \left(\tilde{K}^* f + \partial_t (\mu_1 \nabla_\Omega \cdot \mathbf{j}^k + q^k) + \nabla_\Omega (y^k - \mu_2 z^k) + \mu_3 w^k + r^k + P_\delta \cdot \rho^k \right) \qquad \text{(I)} \\ \mathbf{j}^{k+1} &= \left(\mu_1 \nabla_\Omega \left(\nabla_\Omega \cdot \;\right) - \mu_4 \right)^{-1} \left(-\nabla_\Omega \left(\mu_1\, \partial_t \rho^{k+1} + q^k \right) - \mu_4 \nu^k - \eta^k \right) \qquad \text{(II)} \\ z^{k+1} &= S\left(\nabla_\Omega\, \rho^{k+1} + \frac{1}{\mu_2} y^k \,,\, \frac{\alpha}{\mu_2} \left(\int_\Omega |z^k| dx \right)^{p-1} \right) \quad \forall t \in [0,T] \quad \text{(Shrinkage)} \qquad \text{(III)} \\ 0 &= (w^{k+1})^3 + (\frac{r^k}{\mu_3} - \rho^{k+1}) \cdot (w^{k+1})^2 - \frac{\beta}{\mu_3} \left|\nu^k\right|^2 \qquad \text{(Cubic Equation)} \qquad \text{(IV)} \\ \nu^{k+1} &= \frac{\mu_4\, \mathbf{j}^{k+1} - \eta^k}{\frac{\beta}{w^{k+1}} + \mu_4} \,, \qquad \text{(V)}\end{aligned}$$

where $\tilde{K}$ denotes the operator that is defined as $K\rho(\cdot,t)$ in each time step $t \in [0,T]$.

In the following we will take a closer look at each of the primal updates (I)-(V):

In **step (I)** we need to invert $\tilde{K}^*\tilde{K}$, which is a problem in ill-posed inverse problems, in particular if the underlying operator K cannot be diagonalized. However, here we make use of the preconditioning in the inexact Uzawa algorithm, see P_δ in (I). By choosing

$$P_\delta = \frac{1}{\delta} - \tilde{K}^*\tilde{K} \ ,$$

we get a simplified update for the primal variable ρ without the inversion of $\tilde{K}^*\tilde{K}$:

$$\begin{aligned} \rho^{k+1} \ = \ & \left(\frac{1}{\delta} + \mu_3 - \mu_1\partial_t^2 - \mu_2\Delta_\Omega\right)^{-1} \\ & \left(\tilde{K}^* f + \ \partial_t(\mu_1\nabla_\Omega \cdot \mathbf{j}^k + q^k) + \nabla_\Omega(y^k - \mu_2 z^k) + \mu_3 w^k + r^k + P_\delta \cdot \rho^k\right) \end{aligned} \tag{I}$$

This inversion can be computed efficiently via DCT, since the remaining operators are standard differential operators.

In the second **step (II)** we have a grad-div operator that needs to be inverted. Note that this operator has a non-empty null space. Hence, we have to select the parameters μ_1 and μ_4 carefully. The momentum is a large vector field concerning 4D data. Nevertheless, we can compute this second sub-step efficiently via DCT inversion. Moreover, we can improve the performance of this step by reducing the dimension and by parallelization, see the next subsection.

The variational problem in **step (III)** is a denoising problem with sparsity regularization. In comparison to standard sparsity regularization as presented in Section 4.4.3, we have an additional outer L^p norm here. Nevertheless, we can compute a solution efficiently via a generalized shrinkage formula. For $p = 1$ we obtain a full 4D-TV regularization strategy and a full 4D shrinkage formula, see below.

Since we started with a completely decoupled problem, **step (IV)** results in a simple cubic equation. Similarly to quadratic equations, we can compute a solution of this equation by an explicit formula (including some case differentiations in the implementation).

Step (V) is an explicit update term with pointwise division.

Conclusion: Every step of the algorithm is explicit, DCT inversion or shrinkage. Hence, we obtained a splitting algorithm, where each sub-step can be computed very easily and very efficiently.

Positivity: For simplicity reasons we omitted the positivity constraint due to ρ in the latter presentation. But since we have splitting structure, we can simply extend the algorithm to preserve positivity. The convex positivity constraint can be handled by adding an indicator function $\mathbf{1}_{\rho\geq 0}$ to the objective functional in decoupled problem above. With an additional artificial constraint $\rho_+ = \rho$, we simply have to add the step,

$$\rho_+^{k+1} \;=\; \arg\min_{\rho_+} \left\{ \mathbf{1}_{\rho\geq 0}(\rho_+) \;+\; \frac{1}{\mu_5} \left\| b^k + \rho^{k+1} - \rho_+ \right\|_2^2 \;, \right\}$$

after the ρ update, and we have to add

$$b^{k+1} = b^k + \mu_5(\rho^{k+1} - \rho_+^{k+1})$$

to the dual updates.

Special case: 4D-TV: In the special case of $p = 1$ we obtain a full 4D-TV regularization. Although the analysis we presented relies on the superlinear growth of the TV terms, $p = 1$ seems a reasonable choice. Similarly to the standard spatial case of TV this strategy favors sparsity, i.e. solutions with very large total variation are allowed at some time step, whereas small total variation is favored in the other time steps.

$$\begin{aligned} \rho^{k+1} \;&=\; \left(\tilde{K}^*\tilde{K} + \mu_3 - \mu_1\partial_t^2 - \mu_2\Delta_Q + P_\delta\right)^{-1} && \text{(I)} \\ &\quad \left(\tilde{K}^* f + \;\partial_t(\mu_1 \nabla_Q \cdot \mathbf{j}^k + q^k) + \nabla_Q \cdot (y^k - \mu_2 z^k) + \mu_3 w^k + r^k + P\cdot\rho^k\right) \\ z^{k+1} \;&=\; S(\nabla_Q\, \rho^{k+1} + \frac{1}{\mu_2} y^k \,,\, \frac{\alpha}{\mu_2}) && \text{(Shrinkage)} \quad \text{(III)} \end{aligned}$$

10.5.2 DCT and CUDA and Parallelization

Every step of the presented algorithm is explicit or can be realized via a DCT inversion or shrinkage, i.e. simple thresholding. Hence, every step is very efficient and needs only a minimal amount of memory in each step. Since the system is completely decoupled the resulting sub-steps are highly parallelizable.

As indicated in the previous subsection, the most expensive step resulting in this splitting scheme is the DCT inversion of the grad-div operator in the update of the momentum $\mathbf{j}$ in step (II) of the primal updates. However, we can apply two additional techniques to achieve further efficiency improvements. On the one hand, we can apply a matrix identity, the so-called Woodbury identity, cf. [140],

$$(A + UCV)^{-1} \;=\; A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1},$$

with operators A, U, C and V.

In our case we set $A := -\mu_4$, $C := \mu_1$, $U := \nabla_\Omega$, $V := \nabla_\Omega \cdot$, and obtain

$$\begin{aligned}
\mathbf{j}^{k+1} &= \left(\mu_1 \nabla_\Omega \left(\nabla_\Omega \cdot\ \right) - \mu_4\right)^{-1}(g) \\
&= \left(-\mu_4 + \mu_1 \nabla_\Omega \left(\nabla_\Omega \cdot\ \right)\right)^{-1}(g) \\
&= -\frac{1}{\mu_4} g + \frac{1}{\mu_4} \nabla_\Omega \left[\left(\frac{1}{\mu_1} I - \mu_4 \nabla_\Omega \cdot \nabla_\Omega\right)^{-1} \left(-\frac{1}{\mu_4} \nabla_\Omega \cdot\ g\right)\right] ,
\end{aligned}$$

where

$$g := -\nabla_\Omega \left(\mu_1\, \partial_t \rho^{k+1} + q^k\right) - \mu_4 \nu^k - \eta^k \ .$$

Consequently, this operator identity reduces the dimension of the operator that needs to be inverted via DCT significantly. At the end we just have to deal with a (standard) Laplacian.

In addition to this modification, we implemented the DCT via FFT and used the Mex interface in Matlab to communicate with the CUDA FFT library in C++. Hence, for large data sets we can parallelize the most expensive step on GPUs.

10.5.3 Results - Denoising

In this subsection we will present results of our joint model in the case of $K = I$, i.e. in the case of denoising. We can easily change between isotropic or anisotropic TV in our shrinkage method in step (II). Here we used anisotropic TV.

10.5.4 Results - Deblurring

In this subsection we illustrate the performance of our splitting algorithm in the case where K is a blurring operator and where the data is additionally affected by noise.

10.5.5 Results - Tomography

In this subsection we present mass conservation results in PET imaging. The given data set is of size $175 \times 175 \times 47 \times 10$.

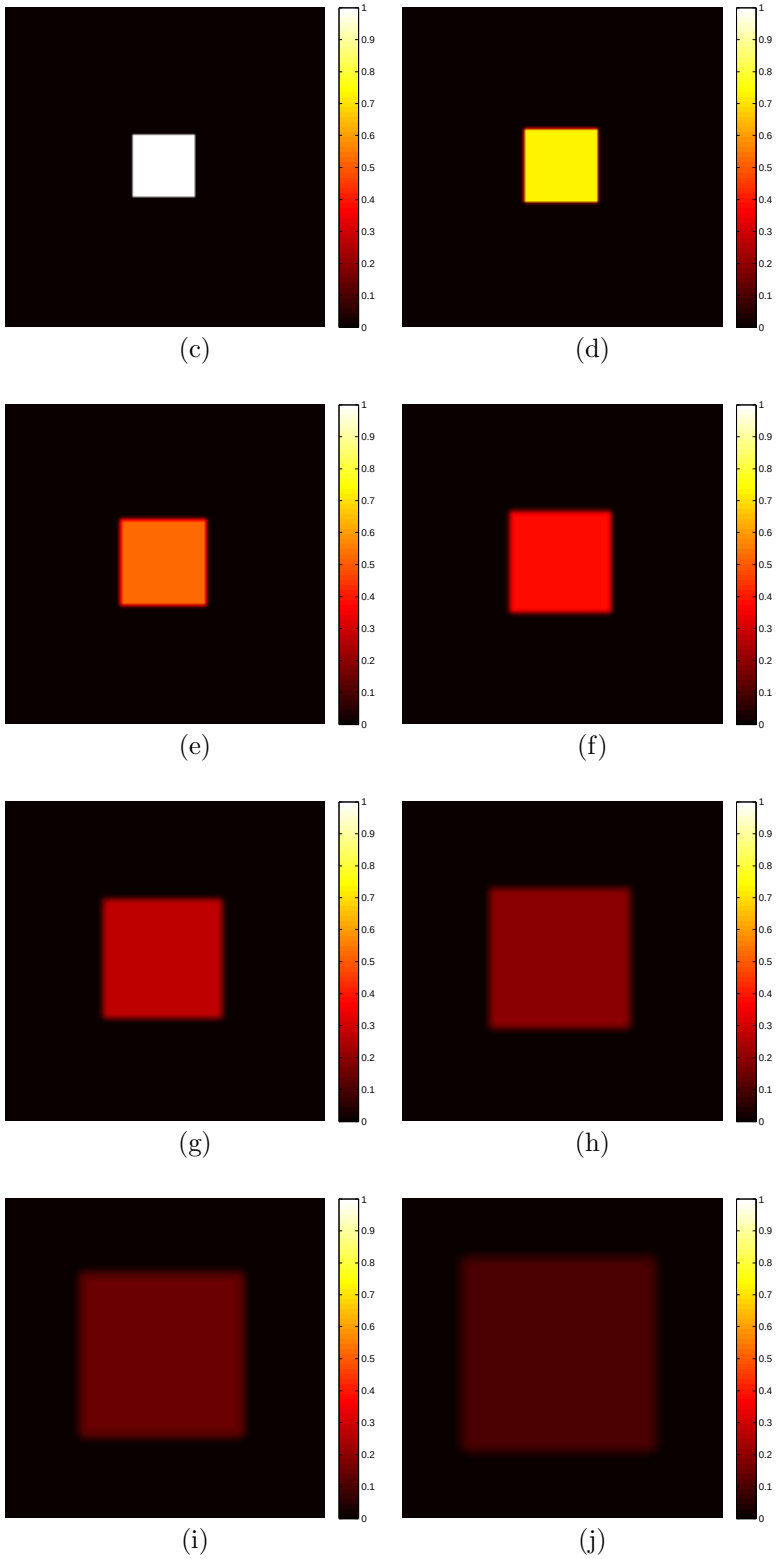


Figure 10.1: **Exact sequence of densities** $\rho(\hat{x}, t)$ for $t = 1...8$ (ground truth)

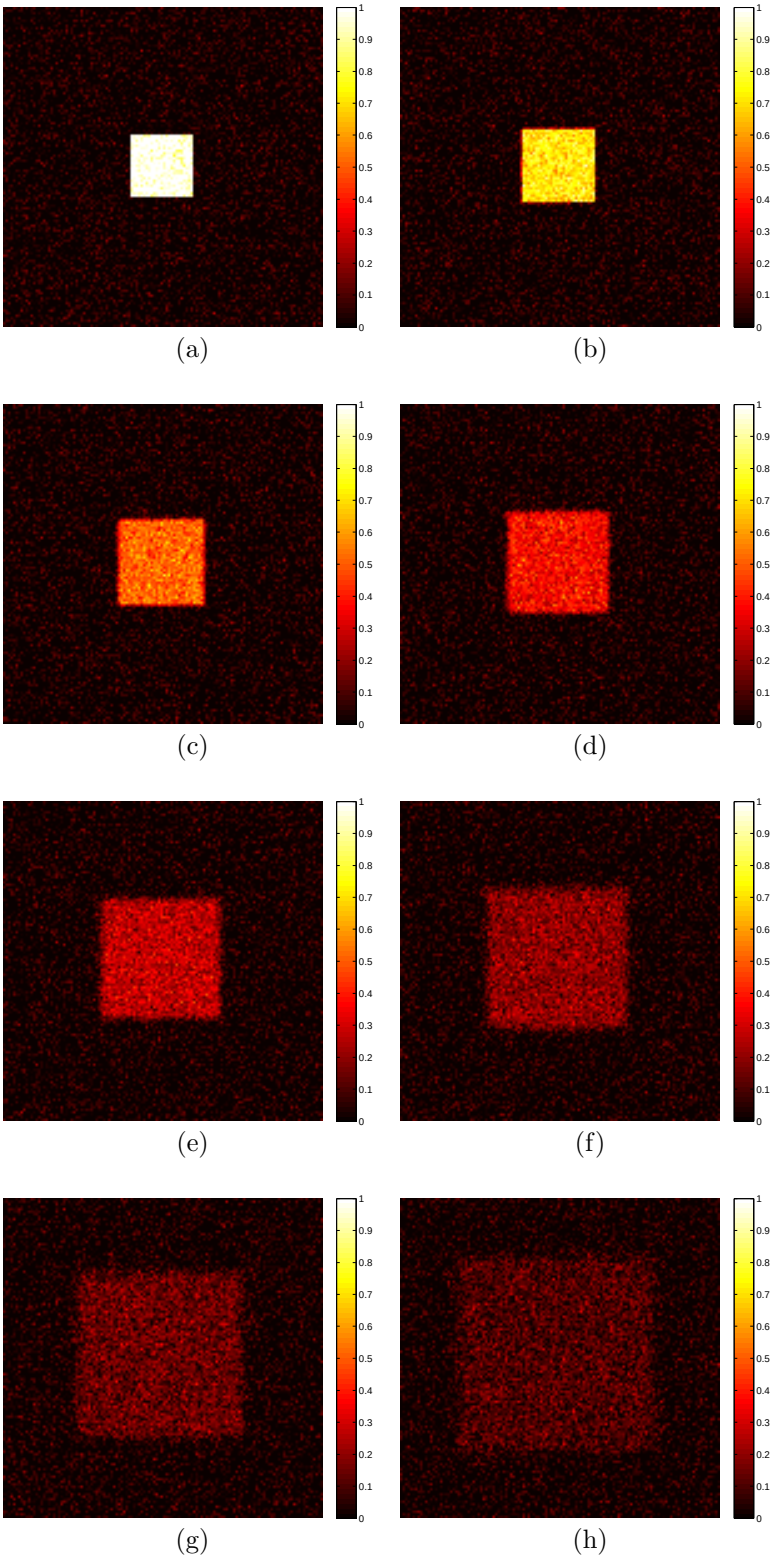


Figure 10.2: Given **noisy sequence of densities** $f(x,t)$ for $t = 1...8$

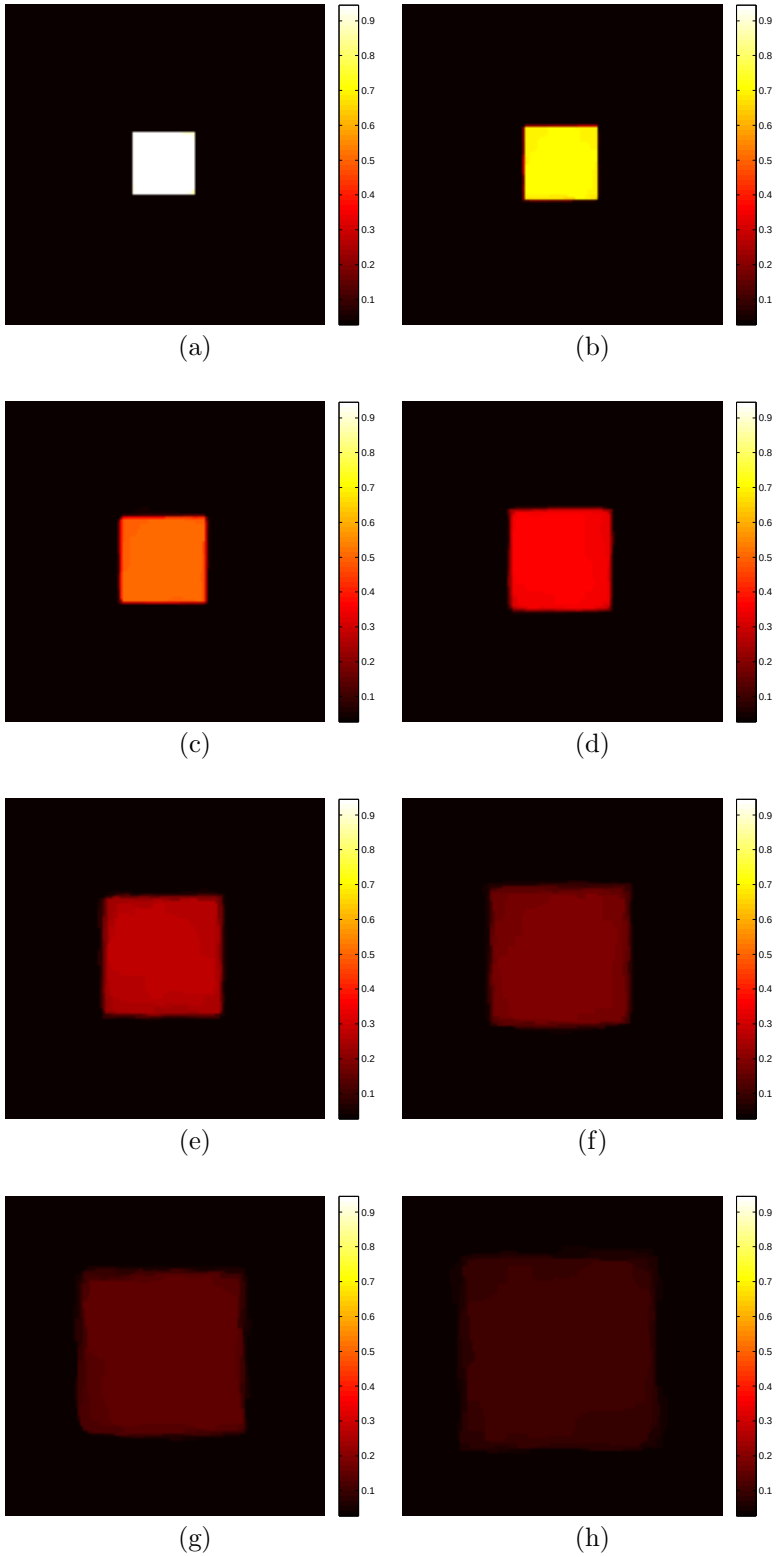


Figure 10.3: **Reconstructed sequence of densities** $\rho(x,t)$ for $t = 1...8$ using the splitting algorithm for reconstruction, total variation and optimal transport.

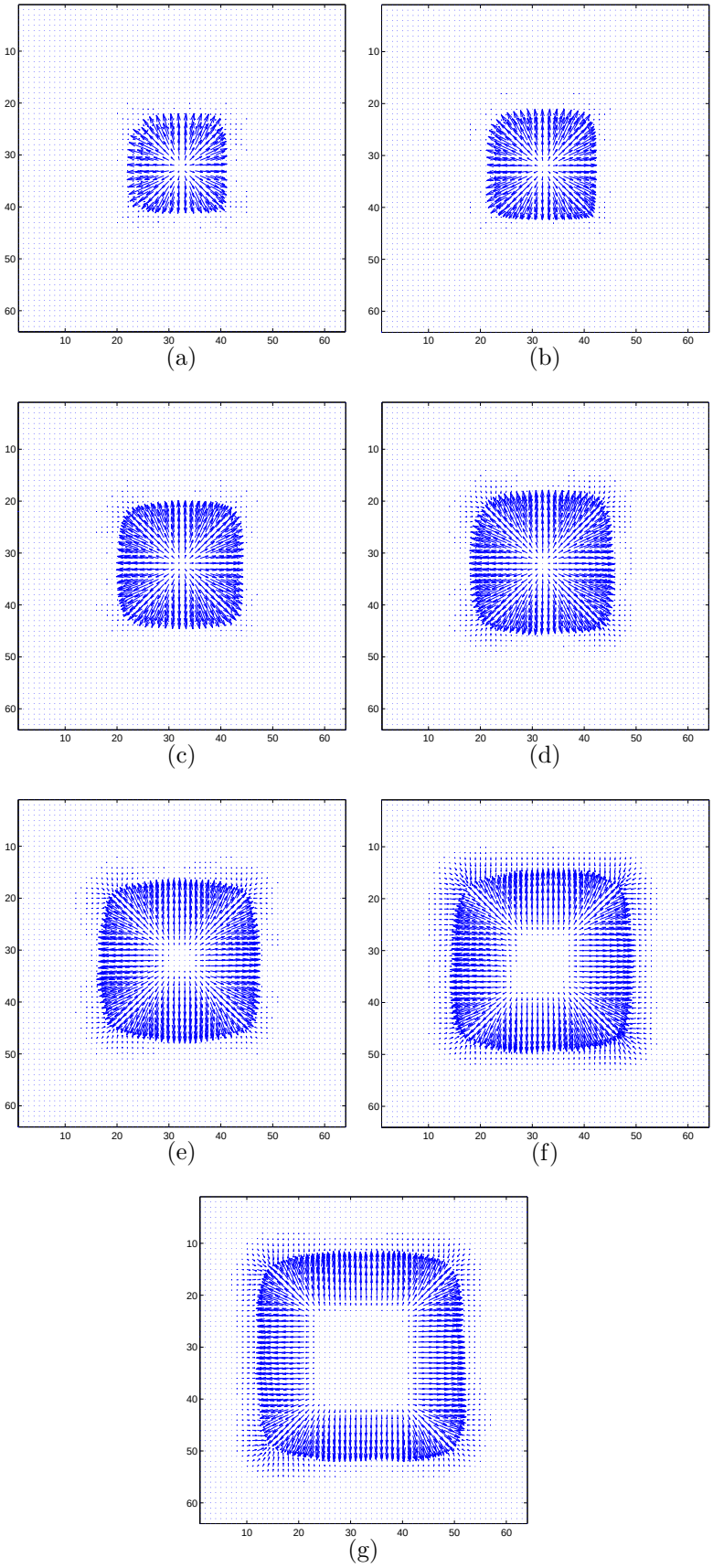


Figure 10.4: Simultaneously computed **motion field** $\mathbf{v}(x,t)$ for $t = 1...7$ using the splitting algorithm including the optimal transport scheme (**vector plot**).

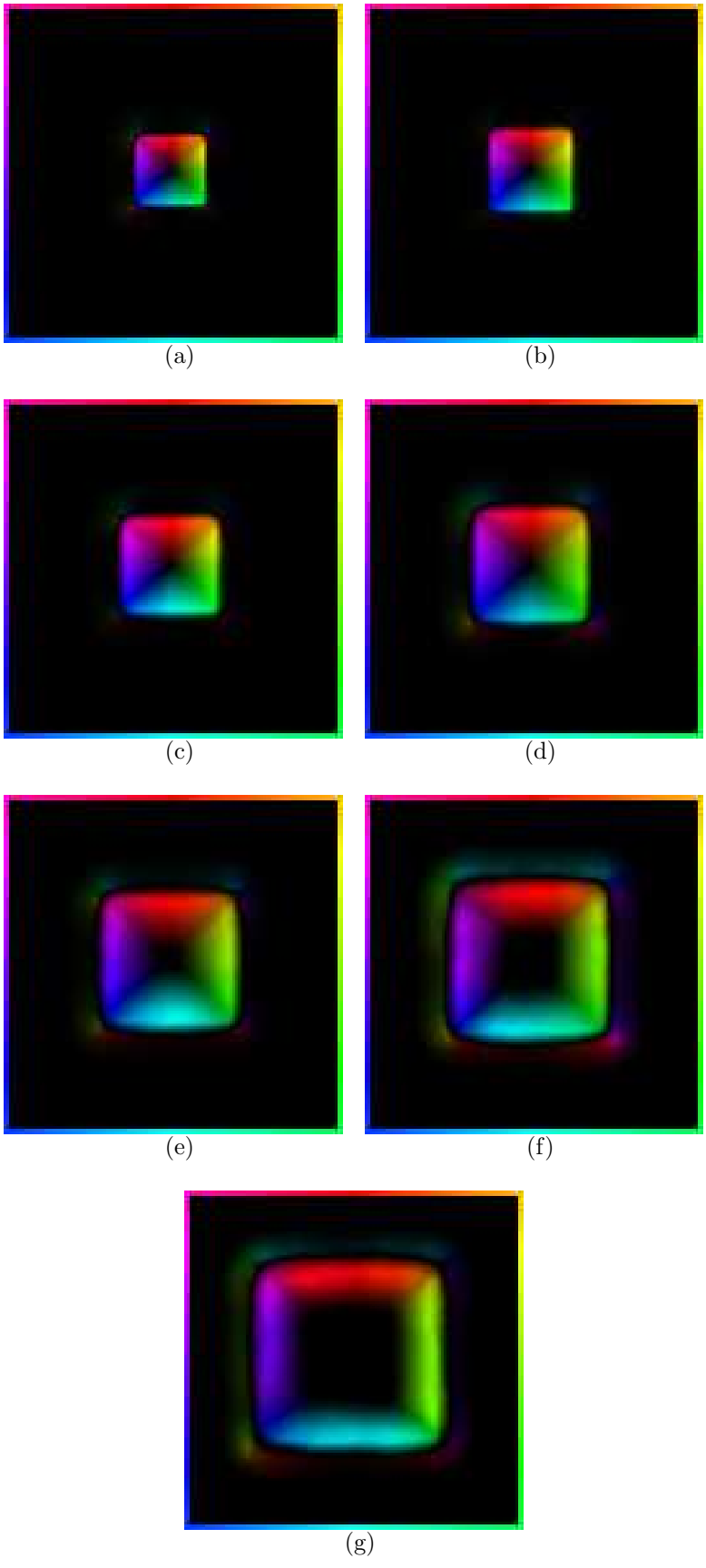


Figure 10.5: Simultaneously computed **motion field** $\mathbf{v}(x,t)$ for $t = 1...7$ using the splitting algorithm including the optimal transport scheme (**colorful orientation plot**).

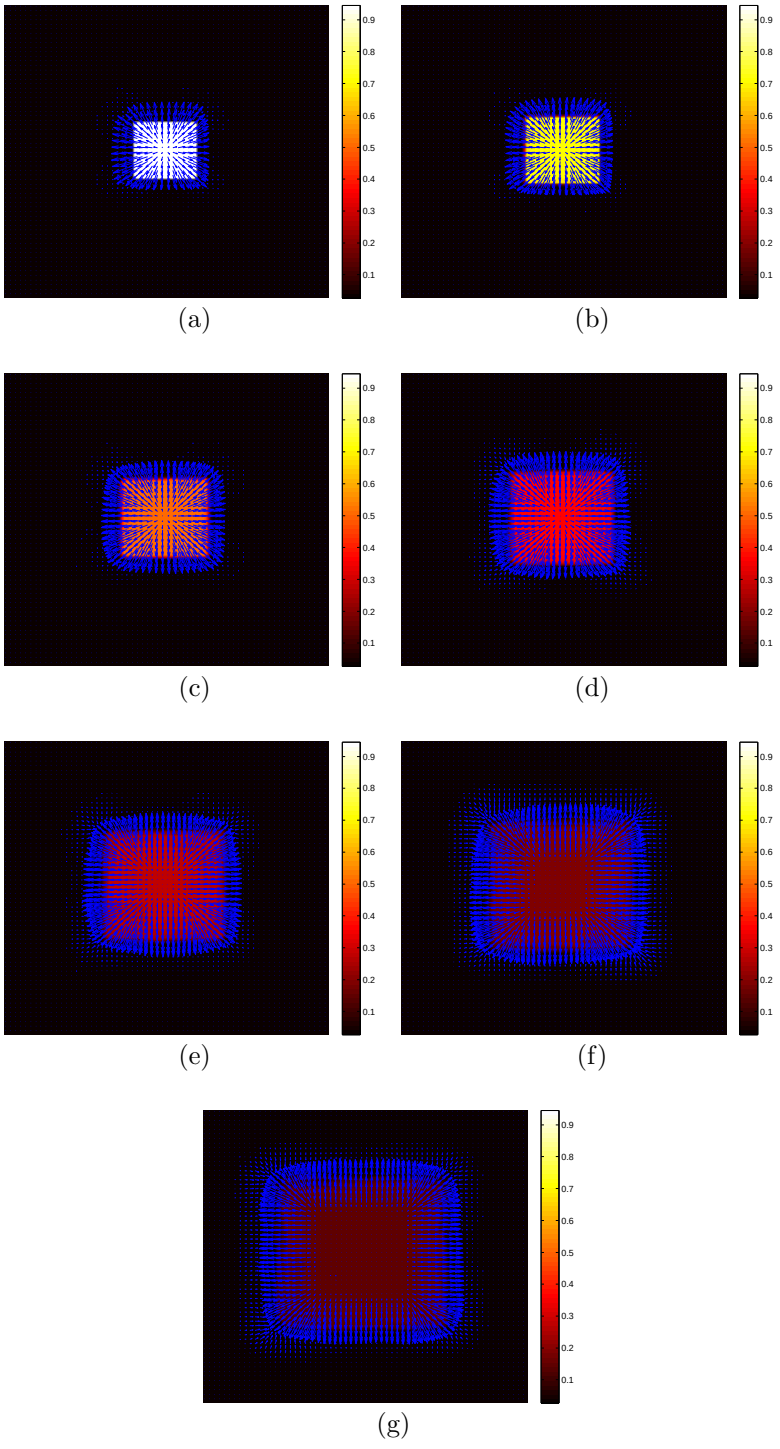


Figure 10.6: Simultaneously computed **sequence of densities** $\rho(x,t)$ **and motion field** $\mathbf{v}(x,t)$ for $t = 1...7$ in one illustration. The velocity field indicates in which way the reconstructed density is transported between each frame. This is very useful for registration purposes.

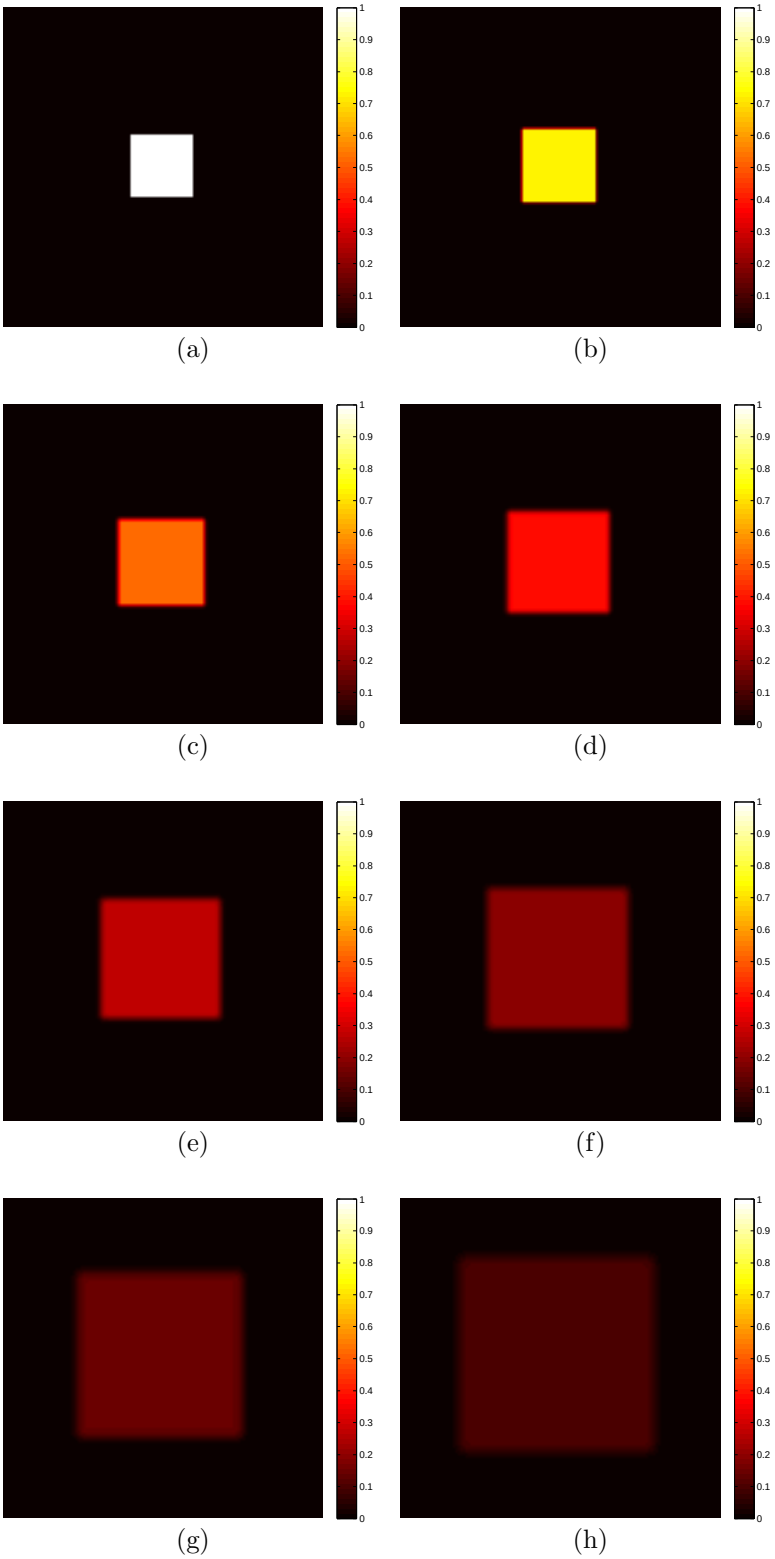


Figure 10.7: **Exact sequence of densities** $\rho(\hat{x}, t)$ for $t = 1...8$ (ground truth)

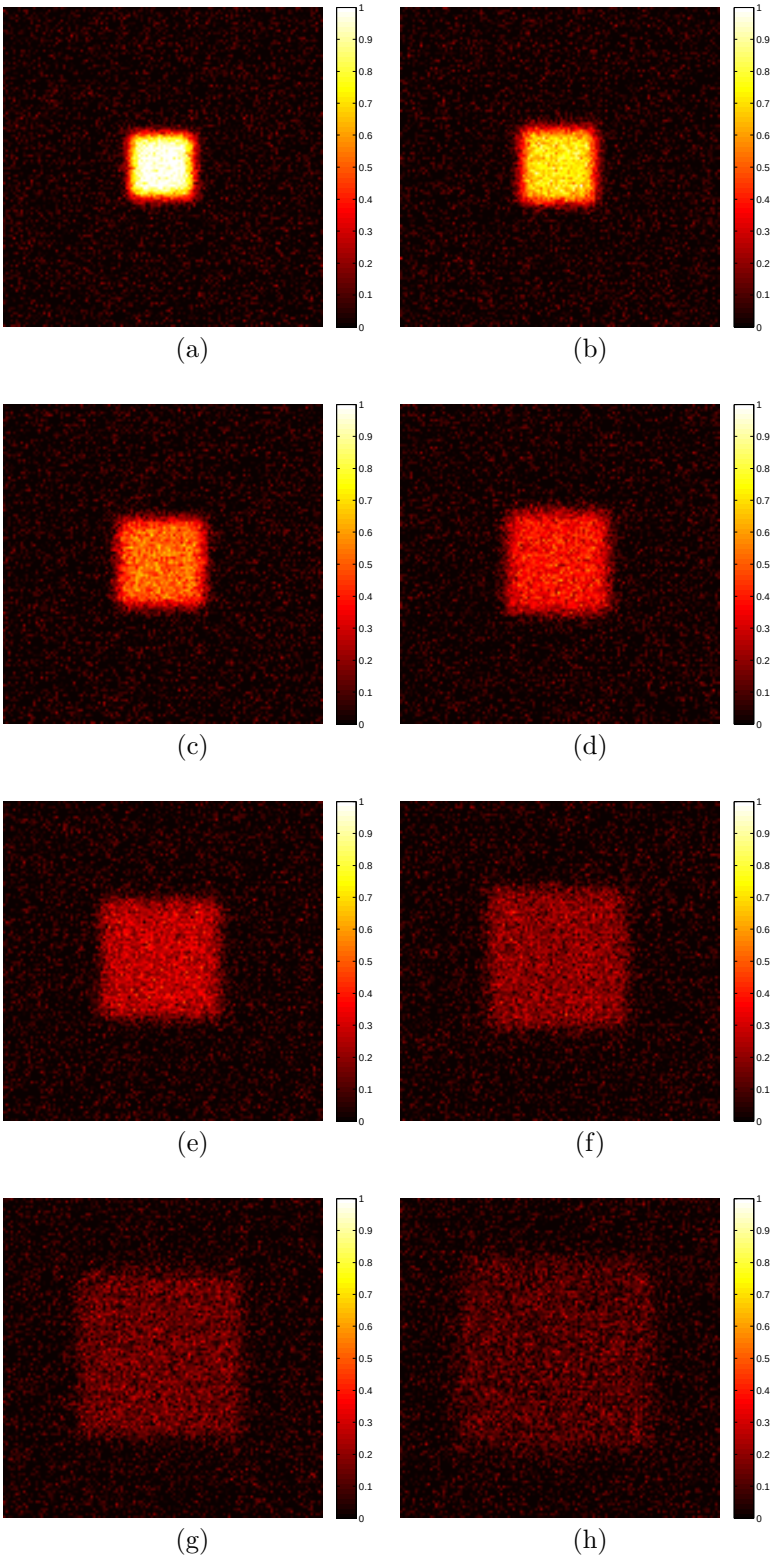


Figure 10.8: Given **blurred and noisy sequence of densities** $f(x,t)$ for $t = 1...8$

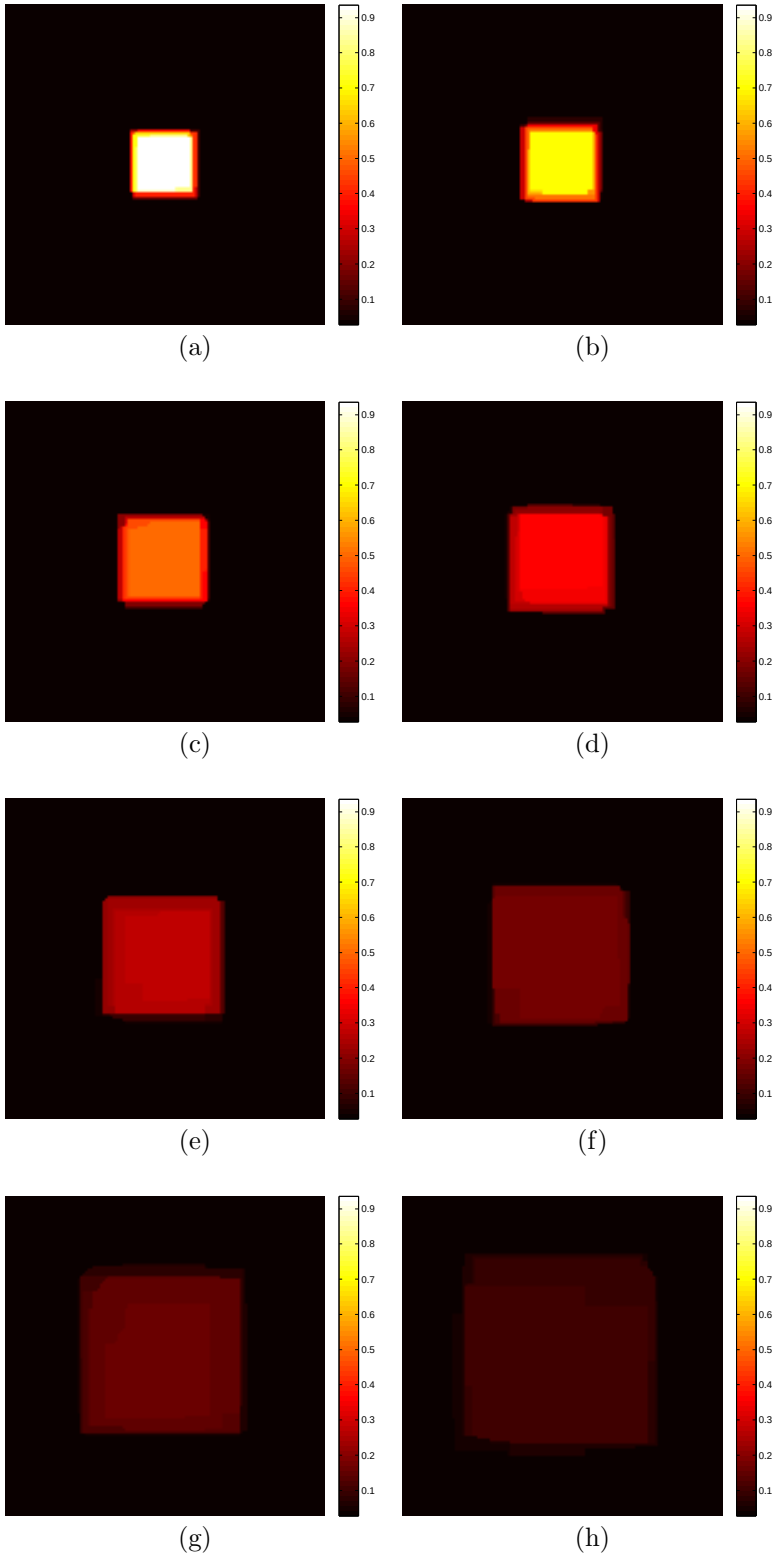


Figure 10.9: **Reconstructed sequence of densities** $\rho(x,t)$ for $t = 1...8$ using the splitting algorithm for reconstruction, total variation and optimal transport.

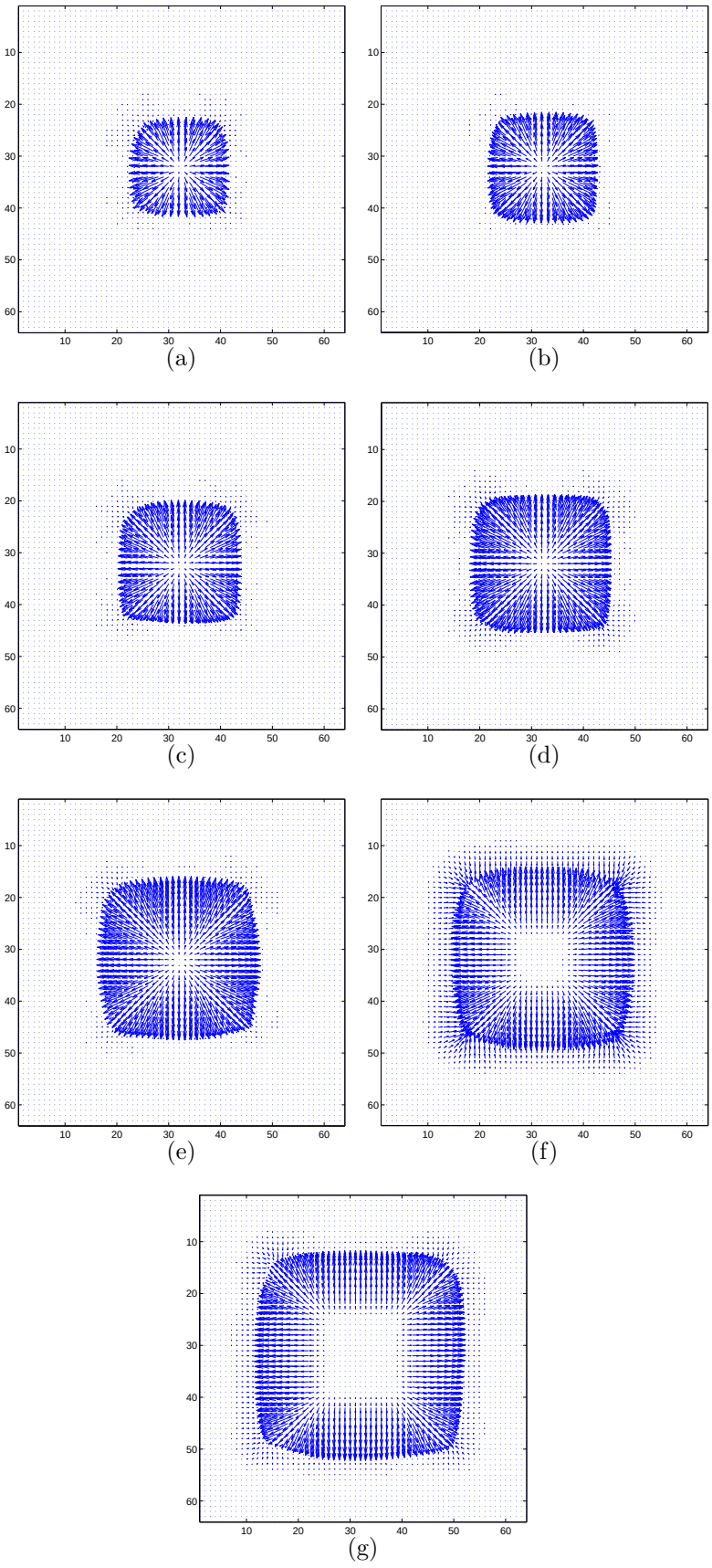


Figure 10.10: Simultaneously computed **motion field** $\mathbf{v}(x,t)$ for $t = 1...7$ using the splitting algorithm including the optimal transport scheme (**vector plot**).

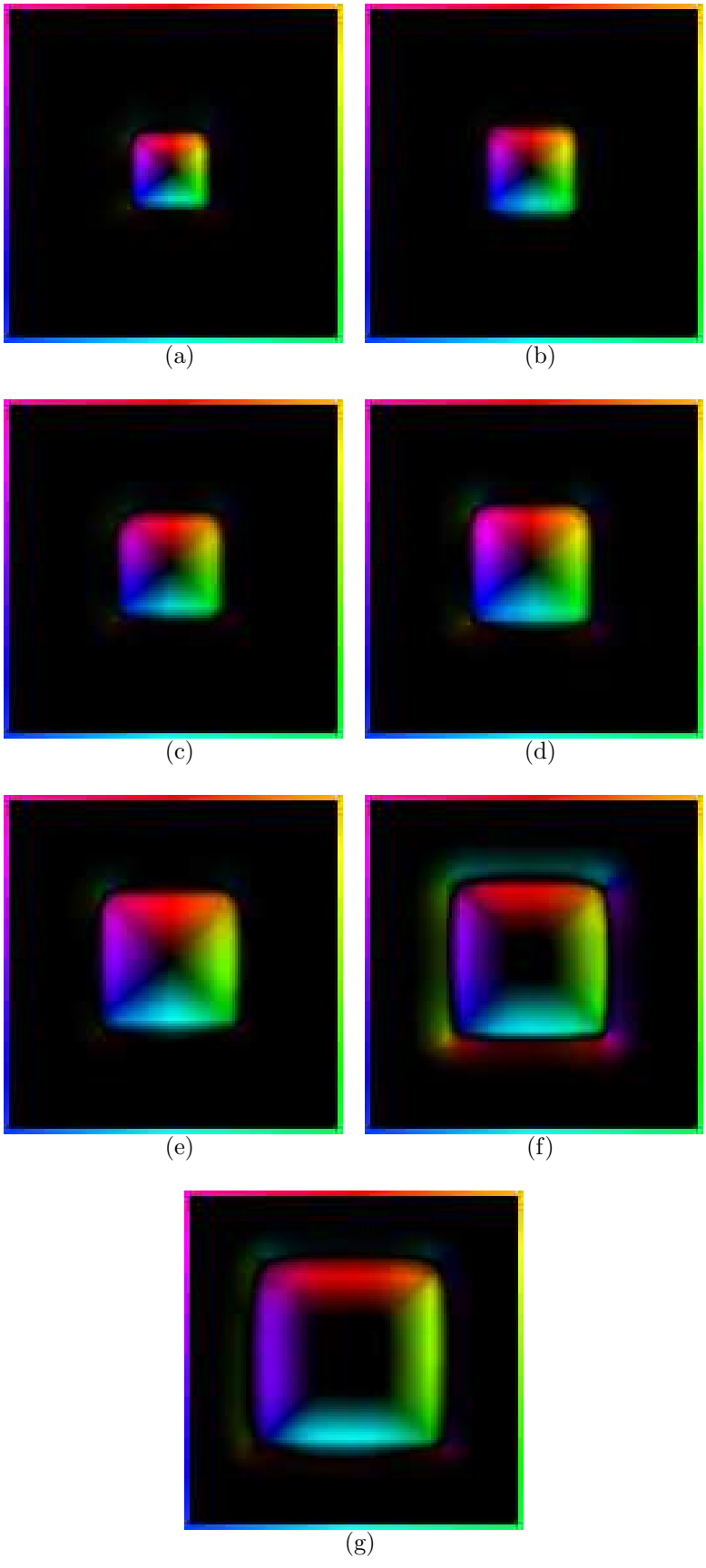


Figure 10.11: Simultaneously computed **motion field** $\mathbf{v}(x,t)$ for $t = 1...7$ using the splitting algorithm including the optimal transport scheme (**colorful orientation plot**).

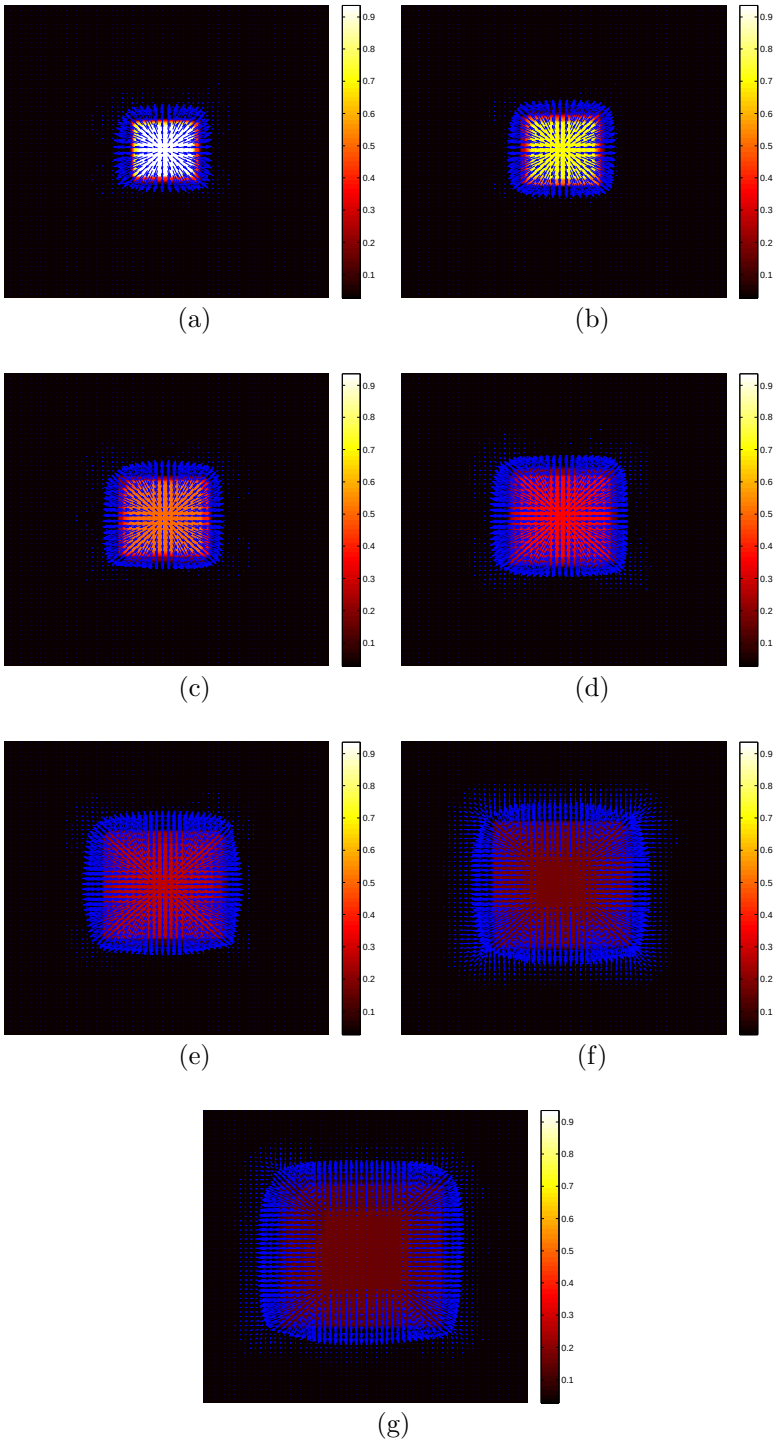
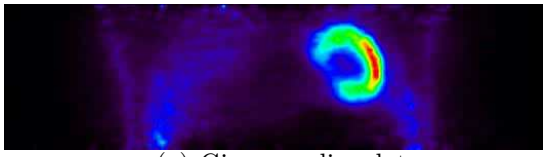
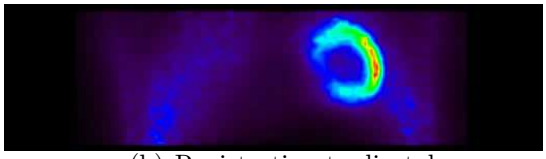


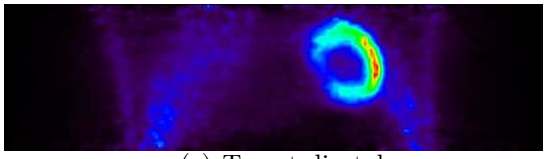
Figure 10.12: Simultaneously computed **sequence of densities** $\rho(x,t)$ **and motion field** $\mathbf{v}(x,t)$ for $t = 1...7$ in one illustration. The velocity field indicates in which way the reconstructed density is transported between each frame. This is very useful for registration purposes.



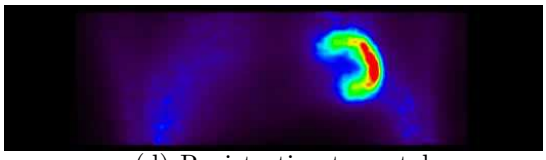
(a) Given cardiac data



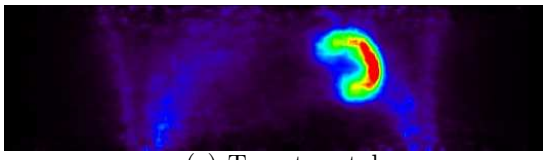
(b) Registration to diastole



(c) Target diastole

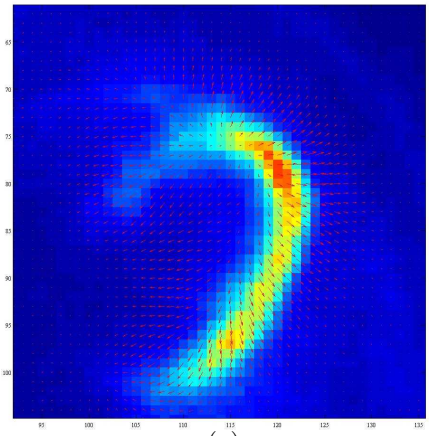


(d) Registration to systole

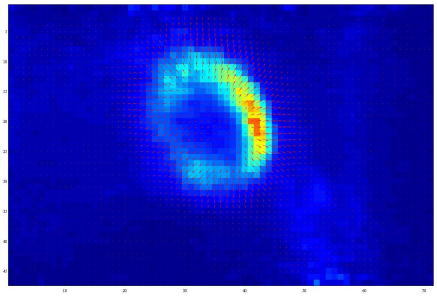


(e) Target systole

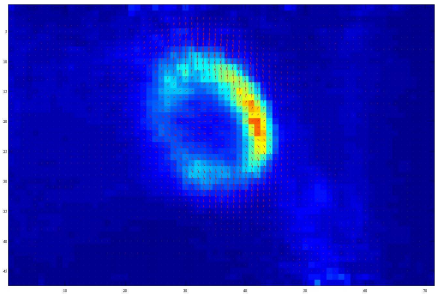
Figure 10.13: Registration results using flow computations via mass conservation.



(a)



(b)



(c)

Figure 10.14: Density ρ with computed motion field $\mathbf{v}$

11

Summary and Open Questions

This thesis contributes to models, analysis and algorithms for 3D static and 4D time-dependent inverse problems in nanoscopy and tomography.

In the first part of this thesis we have considered the static inverse problem

$$Ku(x) \;=\; f(x), \qquad x \in \Omega \;,$$

where only noisy data f were available. For the linear compact operator K we assumed convolution operators in the case of optical nanoscopy and the Radon transform in the case of positron emission tomography (PET). We concentrated on 3D reconstruction models with total variation regularization in the presence of Poisson noise. We proposed an accurate, fast and robust EM-TV algorithm for efficient noise removal and for computing reconstructions facilitating post-segmentation. We studied extensions to simultaneous contrast enhancement via Bregman iterations. From a more general point of view, we answered the question of iterative Bregman regularization for general data fidelities by a new dual Bregman iteration scheme.

Motivated by natural motion effects, e.g. heart beat or breathing in positron emission tomography or living cells in optical nanoscopy, we have considered time-dependent inverse problems,

$$\tilde{K}(\rho(x,t)) \;=\; f(x,t), \qquad (x,t) \in \Omega \times [0,T] \subset \mathbb{R}^3 \times \mathbb{R} \;,$$

in the second part of this thesis. Before we started with a specific modeling for this type of problems, we studied and characterized various models for motion estimation and optimal transport. We pointed out that optical flow techniques only take into account incompressible flows, whereas mass conservation is more general and also takes into account compressible flows. This consideration was useful, in particular for cardiac contraction data in PET imaging. Consequently, we introduced basic

concepts of continuum mechanics and optimal transport problems based on mass conservation. We combined the concepts of motion estimation and optimal transport with the reconstruction ideas of the first part of the thesis, to build a joint 4D model for simultaneous image reconstruction, total variation regularization and optimal transport (including mass conservation).

In Chapter 2 to Chapter 4 we introduced basic analytical as well as numerical concepts, which have been applied throughout the whole thesis. In Chapter 2 we introduced variational methods for inverse problems via MAP estimation and the Bayesian model, discussed different functions spaces including the space BV, introduced total variation and studied basics of differentiability and optimality. Chapter 3 has been dedicated to convex analysis and Bregman distances. We studied the primal and a new dual Bregman iteration for simultaneous contrast enhancement in the case of general variational problems. Moreover, we proved error estimates and revealed interesting properties of Bregman iterations like error forgetting. Splitting methods in convex optimization have been studied in Chapter 4 and have extensively been applied in this thesis. Based on saddle point formulations, augmented Lagrangian methods and inexact Uzawa algorithms, we provided an overview of important splitting techniques in literature, which will be used in this thesis. For example we used forward-backward splitting for solving 3D TV reconstruction problems in Chapter 5, the Split Bregman method for optical flow computations with TV regularization in Chapter 8 or preconditioned inexact Uzawa strategies for our 4D TV reconstruction algorithm in Chapter 10.
In Chapter 5 we studied 3D TV reconstruction problems in the case of Poisson noise. Besides extensions to simultaneous contrast enhancement via Bregman iterations, we presented a detailed analysis of models and the related EM-TV algorithm.
In Chapter 6 the methods on 3D TV reconstruction in the presence of Poisson noise have been applied to 2D and 3D reconstruction problems in optical nanoscopy and positron emission tomography.
After motivating 4D reconstruction via applications, we discussed different motion models based on optical flow in Chapter 8 further in detail. We characterized different data fidelities and different smoothing terms for the flow field. In particular, we combined optical flow techniques with total variation regularization and splitting techniques, and we presented results in high resolution computed tomography (CT).
In Chapter 9 we gave an introduction to basic concepts of continuum mechanics. Based on this, we studied mass conservation and optimal transport problems.
Chapter 10 has been dedicated to the joint 4D image reconstruction model with total variation and optimal transport. We provided a detailed analysis including existence and uniqueness proofs. To overcome the large amount of data we proposed two types of numerical realizations based on preconditioning and splitting techniques

to facilitate efficiency and parallelization. On the one hand we presented a preconditioned Newton-SQP method with integrated line-search. On the other hand, in the case of total variation, we presented a special splitting technique based on inexact Uzawa methods that is highly parallelized and where each of the resulting sub-steps only consists of shrinkage formulas or DCT inversions, which in addition could be parallelized on GPUs.

To summarize, in this thesis we provided several models and algorithms in 3D and 4D image reconstruction including motion estimation, a detailed analysis for different models, as well as a wide range of numerical realizations and results in real applications.

In the following we will give some hints on further research ideas, which have been beyond the scope of this thesis: The connection between preconditioned inexact Uzawa algorithms and Bregmanized operator splitting techniques may possibly open some new insights on error estimation.
A combination of our joint 4D reconstruction models with nonlinear physiological models, to obtain 5D reconstructions, would be an interesting field of future research.
The dual Bregman iteration we proposed in this thesis is a promising approach with nice properties. A further investigation of these properties and an interpretation in terms of optimization would be very interesting.
Total variation regularization and extensions became a quasi standard for static image reconstruction problems and static inverse problems over the past years (resp. decades). However, time-dependent TV-based or, more general, time-dependent sparsity-based reconstruction problems, their analysis and corresponding algorithms have not been considered in detail in literature. This is another interesting field of future research.

Bibliography

[1] R. Acar and C.R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems **10** (1994), 1217–1229. 37, 106, 114, 122, 129, 130, 139

[2] L. Alvarez, J. Esclarín, M. Lefébure, and J. Sánchez, *A PDE model for computing the optical flow*, Proceedings of XVI Congreso de Ecuaciones Diferenciales y Aplicaciones, CEDYA XVI (Las Palmas de Gran Canaria, Spain), 1999, pp. 1349–1356. 188

[3] L. Ambrosio, L. A. Caffarelli, Y. Brenier, G. Buttazzo, and C. Villani, *Optimal Transportation and Applications*, Lecture Notes in Mathematics, Fondazione C.I.M.E., Firenze, no. 1813, Springer, Berlin, Heidelberg, New York, 2001. 213

[4] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, Oxford, New York, 2000. 136

[5] K. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-Linear Programming*, With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow., Stanford Mathematical Studies in the Social Sciences, Stanford University Press, Stanford, Calif., vol. II, 1958. 83, 86, 87

[6] G. Aubert and J. F. Aujol, *A variational approach to remove multiplicative noise*, SIAM J. Appl. Math. **68** (2008), 925–946. 36

[7] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Springer, Berlin, 2002. 48, 122, 124, 126, 177

[8] J.-P. Aubin, *Optima and Equilibria. An Introduction to Nonlinear Analysis*, Graduate Texts in Mathematics, vol. 140, Springer-Verlag, Berlin, 1993. 51

[9] J.-F. Aujol, *Some first-order algorithms for total variation based image restoration*, J. Math. Imaging Vis. **34** (2009), 307–327. 94, 143

[10] D.L. Bailey, D.W. Townsend, P.E. Valk, and M.N. Maisey, *Positron Emission Tomography: Basic Sciences*, Springer, London, 2005. 20, 109, 154

[11] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, *Clustering with Bregman divergences*, Journal of Machine Learning Research **6** (2005), 1705–1749. 63

[12] V. Barbu and T. Precupanu, *Convexity and Optimization in Banach Spaces*, Mathematics and its Applications, Editura Academiei, Bucharest, revised edition, Translated from the Romanian., 1978. 51

[13] J.M. Bardsley, *An efficient computational method for total variation-penalized poisson likelihood estimation*, Inverse Problems and Imaging **2** (2008), 167–185. 106

[14] ———, *A theoretical framework for the regularization of poisson likelihood estimation problems*, Inverse Problems and Imaging **4** (2010), 11–17. 105

[15] J.M. Bardsley and J. Goldes, *Regularization parameter selection methods for ill-posed poisson maximum likelihood estimation*, Inverse Problems **25** (2009), 095005. 118

[16] J.M. Bardsley and A. Luttman, *Total variation-penalized poisson likelihood estimation for ill-posed problems*, Adv. Comput. Math. **31** (2009), 35–59. 105, 106, 130

[17] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numerische Mathematik **84** (2000), 375–393. 27, 172, 212, 215, 218

[18] J.-D. Benamou, Y. Brenier, and K. Guittet, *The Monge-Kantorovitch mass transfer and its computational fluid mechanics formulation*, International Journal for Numerical Methods in Fluids **40** (2002), no. 1-2, 21–30. 27, 172, 212

[19] M. Benning, T. Kösters, F. Wübbeling, K. Schäfers, and M. Burger, *A nonlinear variational method for improved quantification of myocardial blood flow using dynamic $H_2{}^{15}O$ PET*, Nuclear Science Symposium Conference Record, 2008, pp. 4472–4477. 157

[20] M. Benzi, G.H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numerica, Cambridge University Press **14** (2005), 1–137. 83, 86

[21] B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf, *Cartoon extraction based on anisotropic image classification*, Vision, Modeling, and Visualization Proceedings, 2006, pp. 293–300. 144

[22] M. Bertero, H. Lanteri, and L. Zanni, *Iterative image reconstruction: a point of view*, Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT) (Y. Censor, M. Jiang, and A.K. Louis, eds.), Publications of the Scuola Normale, CRM series, vol. 7, 2008, pp. 37–63. 36, 37, 104, 105, 107, 108, 109, 110

[23] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization - Theory and Examples*, Springer, Berlin, 2000. 51, 57

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, 2008. 51, 52, 54, 57

[25] J. H. Bramble, J. E. Pasciak, and A. T. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle point problems*, SIAM J. Numer. Anal. **34** (1997), 1072–1092. 87

[26] K. Bredies, *A forward-backward splitting algorithm for the minimization of non-smooth convex functionals in banach space*, Inverse Problems **25** (2009), 015005. 106, 119

[27] L.M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. and Math. Phys. **7** (1967), 200–217. 60, 64, 145

[28] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr, *A multigrid platform for real-time motion computation with discontinuity-preserving variational methods*, International Journal of Computer Vision **70** (2006), no. 3, 257–277. 185

[29] A. Bruhn, J. Weickert, and C. Schnörr, *Combining the advantages of local and global optic flow methods*, German Pattern Recognition Symposium, 2002, pp. 454–462. 186

[30] C. Brune, H. Maurer, and M. Wagner, *Detection of intensity and motion edges within optical flow via multidimensional control*, SIAM J. Imag. Sci. **2** (2009), no. 4, 1190–1210, Doi: 10.1137/080725064.

[31] C. Brune, A. Sawatzky, and M. Burger, *Bregman-EM-TV methods with application to optical nanoscopy*, Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision (X.-C. Tai et al, ed.), LNCS, vol. 5567, Springer, April 2009, pp. 235–246.

[32] ———, *Primal and dual Bregman methods with application to optical nanoscopy*, International Journal of Computer Vision, Springer Netherlands **0** (2010), 0–0, ISSN: 0920-5691 (Print) 1573-1405 (Online), Doi: 10.1007/s11263-010-0339-5.

[33] M. Burger, K. Frick, S. Osher, and O. Scherzer, *Inverse total variation flow*, Multiscale Model. Simul. **6** (2007), 366–395. 70, 74, 107, 145, 146, 147, 149

[34] M. Burger, G. Gilboa, S. Osher, and J. Xu, *Nonlinear inverse scale space methods*, Commun. Math. Sci. **4** (2006), 179–212. 70, 74, 107, 145, 147, 149

[35] M. Burger and W. Mühlhuber, *Iterative regularization of parameter identification problems by sqp methods*, Inverse Problems **18** (2002), 943–970. 83

[36] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems **20** (2004), 1411–1421. 63, 64, 74

[37] M. Burger, E. Resmerita, and L. He, *Error estimation for bregman iterations and inverse scale space methods*, Computing **81** (2007), 109–135. 70

[38] A. Cambini and L. Martein, *Generalized Convexity and Optimization - Theory and Applications*, Lecture Notes in Economics and Mathematical Systems, vol. 616, Springer, Berlin, Heidelberg, 2009. 54

[39] Z.-H. Cao, *Fast Uzawa algorithm for generalized saddle point problems*, Applied Numerical Methods **46** (2003), 157–171.

[40] J. A. Carrillo, A. Jüngel, P.A. Markowich, G. Toscani, and A. Unterreiter, *Entropy dissipation methods for degenerate parabolic problems and generalized sobolev inequalities*, Monatshefte für Mathematik, Springer Wien **133** (2001), no. 1, 1–82. 75

[41] J. A. Carrillo, R. J. McCann, and C. Villani, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Springer, Berlin, Heidelberg **179** (2006), no. 2, 217–263. 75

[42] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis. **20** (2004), 89–97. 106, 116, 143, 144

[43] ———, *Total variation minimization and a class of binary MRF models*, Energy Minimization Methods in Computer Vision and Pattern Recognition (A. Rangarajan, B. C. Vemuri, and A. L. Yuille, eds.), EMMCVPR, vol. LNCS 3757, Springer, 2005, pp. 136–152. 94, 143

[44] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, *Two deterministic half-quadratic regularization algorithms for computed imaging*, Pro-

ceedings of the IEEE International Conference on Image Processing (Austin, Texas), vol. 2, 1994, pp. 168–172. 188

[45] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim. **3** (1993), no. 3, 538–543. 60, 64

[46] I. Cohen, *Nonlinear variational method for optical flow computation*, Proceedings of the Scandinavian Conference on Image Analysis (Göteborg, Schweden), 1993, pp. 523–530. 190

[47] P. L. Combettes, *Solving monotone inclusions cia compositions of nonexpansive averaged operators*, Optimization **53** (2004), no. 5-6, 475–504. 96

[48] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul. **4** (2005), 1168–1200. 119

[49] P.L. Combettes and J.-C. Pesquet, *A proximal decomposition method for solving convex variational inverse problems*, Inverse Problems **24** (2008), 065014. 119

[50] I. Csiszar, *Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems*, Ann. Statist. **19** (1991), 2032–2066. 121

[51] M. Cui, *A sufficient condition for the convergence of the inexact Uzawa algorithm for saddle point problems*, Journal of Computational and Applied Mathematics **139** (2002), 189–196.

[52] R. Dautray and J.-L. Lions, *Mathematical analysis and numerical methods for science and technology, volume 3*, Springer, 2000. 197

[53] M. Dawood, C. Brune, X. Jiang, M. Burger, O. Schober, M. Schäfers, and K. Schäfers, *A continuity equation based optical flow method for cardiac motion correction in 3D PET data*, Tech. report, submitted to MICCAI Medical Image Computing and Computer Assisted Intervention, Mai 2010.

[54] M. Dawood, F. Büther, X. Jiang, and K. P. Schäfers, *Respiratory motion correction in 3D PET data with advanced optical flow algorithms*, IEEE Trans Med Imaging **27** (2008), 1164–1175. 26, 170

[55] A.P. Dempster, N.M. Laird, and D.B. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*, Journal of the Royal Statistical Society, Series B **39** (1977), no. 1, 1–38. 105, 111, 112

[56] R. Deriche, P. Kornprobst, and G. Aubert, *Optical flow estimation while preserving its discontinuities: A variational approach*, In Proceedings of the Asian Conference on Computer Vision, ACCV'95 (Singapore), 1995. 190

[57] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivio-Marin, and J. Zerubia, *3D Microscopy Deconvolution using Richardson-Lucy Algorithm with Total Variation Regularization*, Tech. Report 5272, Institut National de Recherche en Informatique et en Automatique, 2004. 106, 109

[58] J. Douglas and H.H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Americ. Math. Soc. **82** (1956), 421–439. 95, 119

[59] J. Eckstein, *Splitting methods for monotone operators with application to parallel optimization*, Ph.D. thesis, MIT, 1989.

[60] J. Eckstein and D. P. Bertsekas, *On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Prog. **55** (1992), 293–318. 95

[61] P.P.B. Eggermont, *Maximum entropy regularization for fredholm integral equations of the first kind*, SIAM J. Math. Anal. **24** (1993), 1557–1576. 121, 122

[62] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, Studies in Mathematics and Its Applications, vol. 1, North-Holland Publishing Company - Amsterdam/Oxford, 1976. 51, 56, 71, 73, 84, 114, 143

[63] H. C. Elman and G.H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal. **31** (1994), 1645–1661. 54, 87

[64] J. Elstrodt, *Maß- und Integrationstheorie*, Springer; Berlin etc., 1996. 44

[65] H.W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publisher - Dordrecht, 1996. 21, 110

[66] Y. E. Erdi, Sadek A. N., T. Pan, A. Pevsner, K. E. Rosenzweig, G. Mageras, E. D. Yorke, H. Schroder, W. Hsiao, O. D. Squire, P. Vernon, J. B. Ashman, H. Mostafavi, S. M. Larson, and J. L. Humm, *The CT motion quantitation of lung lesions and its impact on PET-measured SUVs*, J Nucl Med **45** (2004), no. 8, 1287–1292. 25, 170

[67] S. Esedoḡlu and S.J. Osher, *Decomposition of images by the anisotropic rudin-osher-fatemi model*, Comm. Pure Appl. Math. **57** (2004), 1609–1626. 144

[68] E. Esser, *Applications of lagrangian-based alternating direction methods and connections to split bregman*, Tech. report, UCLA CAM Report 09-31, 2009. 87, 89, 97

[69] E. Esser, X. Zhang, and T. Chan, *A general framework for a class of first order primal-dual algorithms for tv minimization*, Tech. report, UCLA CAM Report 09-67, 2009. 87, 89, 97

[70] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992. 106, 114, 122

[71] M. Fortin and A. Fortin, *A generalization of Uzawa´s algorithm for the solution of the Navier-Stokes equations*, Communications in Applied Numerical Methods **1** (1985), 205–208.

[72] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, Elsevier Science Publishers B.V., Amsterdam, Studies in Mathematics and its Applications, 1983. 87

[73] M.A. Freitag, N.K. Nichols, and C.J. Budd, *Resolution of sharp fronts in the presence of model error in variational data assimilation*, Preprint, Bath Institute For Complex Systems, University of Bath, January 2010. 221

[74] K. Frick, *The augmented Lagrangian method and associated evolution equations*, Ph.D. thesis, University of Innsbruck, 2008. 88

[75] K. Frick and O. Scherzer, *Regularization of ill-posed linear equations by the non-stationary augmented lagrangian method*, Tech. report, J. Integral Equations Appl., to appear, 2009. 88

[76] D. Gabay, *Applications of the method of multipliers to variational inequalities*, In: Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems (M. Fortin and R. Glowinski, eds.), vol. 15, Elsevier Science Publishers B.V., Amsterdam, Studies in Mathematics and its Applications, 1983, pp. 299–331. 94, 95, 132

[77] A. L. Garcia, *Numerical Methods for Physics*, Prentice Hall, 1994.

[78] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell. **6** (1984), 721–741. 35

[79] S. Geman and D. E. McClure, *Bayesian image analysis: an application to single photon emission tomography*, Statistical Computation Section, American Statistical Association, Washington, 1985, pp. 12–18. 35

[80] B. A. Gilland, D. R.and Mair and J. G. Parker, *Motion estimation for cardiac emission tomography by optical flow methods*, Phys. Med. Biol. **53** (2008), 2991–3006. 26, 170

[81] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Monographs in Mathematics, vol. 80, Birkhäuser - Boston/Basel/Stuttgart, 1984. 114, 122

[82] R. Glowinski, *Numerical methods for nonlinear variational problems*, Springer Series in Computational Physics, Springer, New York, 1984. 83

[83] Roland Glowinski and Patrick Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM, Studies in Applied Mathematics, 1989. 84, 87, 89, 95

[84] T. Goldstein and S. Osher, *The split bregman method for L1-regularized problems*, SIAM J. Imaging Sci. **2** (2009), 323–343. 97, 119

[85] C. Greif and D. Schötzau, *Preconditioners for saddle point linear systems with higly singular (1,1) blocks*, Electronic Transactions on Numerical Analysis **22** (2006), 114–121. 228

[86] C.W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Vieweg Verlag - Braunschweig/Wiesbaden, 1993. 110

[87] C. Gräser and R. Kornhuber, *On preconditioned Uzawa-type iterations for a saddle point problem with inequality constraints*, Lecture Notes in Computational Science and Engineering, vol. 55, ch. Part I, pp. 91–102, Springer, Berlin, Heidelberg, 2007.

[88] E. Haber and U. M. Ascher, *Preconditioned all-at-once methods for large, sparse parameter estimation problems*, Inverse Problems **17** (2001), 1847–1864. 83

[89] E. Haber, U. M. Ascher, and D. Oldenburg, *On optimization techniques for solving nonlinear inverse problems*, Inverse Problems **16** (2000), 1263–1280. 83

[90] E. Haber, R. Horesh, and J. Modersitzki, *Numerical optimization for constrained image registration*, Numerical Linear Algebra with Applications **17** (2010), no. 2-3, 343–359. 172

[91] E. Haber and J. Modersitzki, *Numerical methods for volume-preserving image registration*, Inverse Problems **20** (2004), 1621–1638. 83, 172

[92] E. Haber, T. Rehman, and A. Tannenbaum, *An efficient numerical method for the solution of the L2 optimal mass transfer problem*, SIAM J. on Scientific Computing **32** (2010), 197–211. 172, 213

[93] E. T. Hale, W. Yin, and Y. Zhang, *A fixed-point continuation method for l1-regularization with application to compressed sensing*, Tech. report, Rice University CAAM Technical Report TR07-07, 2007. 94

[94] ———, *A fixed-point continuation applied to compressed sensing: implementation and numerical experiments*, Journal of Computational Mathematics **28** (2010), no. 2, 170–194. 94

[95] B. Hein, K. I. Willig, C. A. Wurm, V. Westphal, S. Jakobs, and S. W. Hell, *Stimulated emission depletion nanoscopy of living cells using SNAP-tag fusion proteins*, Biophysical Journal **98** (2010), 158–163. 25, 169

[96] S. Hell and A. Schönle, *Nanoscale resolution in far-field fluorescence microscopy*, Science of Microscopy (P.W. Hawkes and J.C.H. Spence, eds.), Springer, 2006. 18, 153

[97] S. Hell and E.H.K. Stelzer, *Properties of a 4Pi confocal fluorescence microscope*, J. Opt. Soc. Am. A **9** (1992), 2159–2166. 18, 153, 155

[98] S.W. Hell, *Toward fluorescence nanoscopy*, Nat. Biotechnol. **21** (2003), 1347–1355. 109

[99] M. R. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications **4** (1969), 303–320. 87, 88

[100] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms I*, Grundlehren der mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 305, Springer-Verlag, 1993. 51, 112, 123, 144

[101] ———, *Fundamentals of Convex Analysis*, Springer, Berlin, Heidelberg, 2004. 51

[102] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, *A convergence rates result for tikhonov regularization in banach spaces with non-smooth operators*, Inverse Problems **23** (2007), no. 3, 987–1010. 63

[103] B. K. P. Horn and B. Schunck, *Determining optical flow*, Artificial Intelligence **59** (1993), no. 1-2, 81–87. 183

[104] P. J. Huber, *Robust Statistics*, John Wiley and Sons; New York, 1981. 190

[105] A. D. Ioffe and V. M. Tihomirov, *Theorie der Extremalaufgaben*, VEB Deutscher Verlag der Wissenschaften, Berlin, Translated from the Russian by Bernd Luderer, 1979. 51

[106] K. Ito and K. Kunisch, *An active set strategy based on the augmented lagrangian formulation for image restoration*, M2AN Math. Model. Numer. Anal. **33** (1999), no. 1, 1–21. 87

[107] ———, *Lagrange Multiplier Approach to Variational Problems and Applications*, vol. 15, Advances in Design and Control, SIAM, Philadelphia, PA, 2008. 87

[108] A.N. Iusem, *Convergence Analysis for a Multiplicatively Relaxed EM Algorithm*, Mathematical Methods in the Applied Sciences **14** (1991), no. 8, 573–593. 112

[109] E. Jonsson, S.C. Huang, and T. Chan, *Total variation regularization in positron emission tomography*, CAM Report 98-48, UCLA, 1998. 106

[110] R. J. Kittel and et al., *Bruchpilot promotes active zone assembly, Ca2+ channel clustering, and vesicle release*, Science **312** (2006), 1051–1054. 158

[111] K. Kiwiel, *Proximal minimization methods with generalized bregman functions*, SIAM J. Cont. Optim. **35** (1997), 1142–1168. 60, 64

[112] T.A. Klar, S. Jacobs, M. Dyba, A. Egner, and S.W. Hell, *Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission*, PNAS **97** (2000), 8206–8210. 18, 153

[113] A. Kumar, A. Tannenbaum, and G. J. Balas, *Optical flow: a curve evolution approach*, IEEE Transactions on Image Processing **5** (1996), no. 4, 598–610. 190

[114] J.-M. Lasry and P.-L. Lions, *Mean field games*, Japan. J. Math. **2** (2007), 229–260. 222

[115] M. Lauterbach, *Fast STED microscopy*, Ph.D. thesis, University of Göttingen, Germany, 2009. 25, 169

[116] T. Le, R. Chartrand, and T.J. Asaki, *A variational approach to reconstructing images corrupted by poisson noise*, J. Math. Imaging Vis. **27** (2007), 257–263. 106

[117] H. Liao, F. Li, and M.K. Ng, *Selection of regularization parameter in total variation image restoration*, J. Opt. Soc. Am. A **26** (2009), 2311–2320. 118

[118] Y. Lin and Y. Wei, *Fast corrected Uzawa methods for solving symmetric saddle point problems*, CALCOLO, Springer Milan **43** (2006), 1126–5434.

[119] P.L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal. **16** (1979), 964–979. 93, 95, 96, 106, 119, 120

[120] B. D. Lucas, *Generalized image matching by the method of differences*, Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 1984. 186

[121] B. D. Lucas and T. Kanade, *An iterative image registration technique with an application to stereo vision*, Proceedings of the DARPA Image Understanding Workshop, Washington, D.C, 1981, pp. 121–130. 186

[122] L.B. Lucy, *An iterative technique for the rectification of observed distributions*, Astronomical Journal **79** (1974), no. 6, 745–754. 105, 111

[123] B. A. Mair and D. R. Gillard, *Estimation of images and nonrigid deformations in gated emission CT*, IEEE Transactions on Medical Imaging **25** (2006), 1130–1144. 26, 170

[124] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, University Lecture Series, vol. 22, American Mathematical Society, Boston, MA, USA, 2001. 22, 66, 144, 145, 148

[125] H. Miike, A. Nomura, and K. Koga, *Determining image flow from multiple frames based on the continuity equation*, Technology Report Yamaguchi University **4** (1991), no. 5, 387–397.

[126] J. Modersitzki, *Numerical methods for image registration*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2003. 83, 172

[127] ———, *Fair: Flexible algorithms for image registration*, SIAM, Philadelphia, 2009. 83, 172

[128] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l' Académie Royale des Science, Année 1781, avec les Mémoires de Mathématique et de Physique (1781). 27, 172, 210, 212

[129] J. Müller, *Parallel total variation minimization*, Master's thesis, Institute for Computational and Applied Mathematics, University of Münster, 2008. 143

[130] H. H. Nagel and W. Enkelmann, *An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences*, IEEE Trans. Pattern Analysis and Machine Intelligence **8** (1986), no. 5, 565–593. 188

[131] M. Nagorni and S.W. Hell, *4Pi-Confocal Microscopy Provides Three-Dimensional Images of the Microtubule Network with 100- to 150-nm Resolution*, J. Struct. Biol. **123** (1998), 236–247. 109

[132] Y. Nakamoto, B. B. Chin, C. Cohade, M. Osman, M. Tatsumi, and R. L. Wahl, *PET/CT: artifacts caused by bowel motion*, Nucl Med Commun **25** (2004), no. 3, 221–225. 25, 170

[133] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM Monographs on Mathematical Modeling and Computation, 2001. 109, 111, 112, 140

[134] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul. **4** (2005), 460–489. 22, 66, 107, 145, 147, 149, 150

[135] M. M. Osman, C. Cohade, Y Nakamoto, L. T. Marshall, J. P. Leal, and R. L. Wahl, *Clinically significant inaccurate localization of lesions with PET/CT: frequency in 300 patients*, J Nucl Med **44** (2003), no. 2, 240–243. 25, 170

[136] V.Y. Panin, G.L. Zeng, and G.T. Gullberg, *Total variation regulated EM algorithm [SPECT reconstruction]*, IEEE Trans. Nucl. Sci. **46** (1999), 2202–2210. 106

[137] D. Papathanassiou, S. Becker, R. Amir, B. Menéroux, and J.-C. Liehn, *Respiratory motion artifact in the liver dome on FDG PET/CT comparison of attenuation correction with ct and a caesium external source*, Eur J Nucl Med Mol Imaging **32** (2005), 1422–1428. 26, 170

[138] G.B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert spaces*, J. Math. Anal. Appl. **72** (1979), 383–390. 93, 119

[139] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans Pattern Anal Mach Intell **12** (1990), 629–639. 70

[140] K. B. Petersen and M. S. Pedersen, *The matrix cookbook*, oct 2008, Version 20081110. 234

[141] J. Pietschmann, *Long-time behaviour of nonlinear fokker-planck equations*, Master's thesis, Institute for Computational and Applied Mathematics, University of Münster, 2008. 75

[142] L. Popov, *A modification of the Arrow-Hurwicz method for search of saddle points*, Mathematical Notes **28** (1980), no. 5, 845–848. 87

[143] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, Optimization, pp. 283–298, Academic Press London, New York, 1969. 87, 88

[144] F. Qiao, T. Pan, J. W. Clark Jr, and O. R. Mawlawi, *A motion-incorporated reconstruction method for gated pet studies*, Phys. Med. Biol. **51** (2006), 3769–3783. 26, 170

[145] W. Queck, *The convergence factor of preconditioned algorithms of the Arrow-Hurwicz type*, SIAM J. Numer. Anal. **26** (1989), 1016–1030. 87

[146] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems - Volume I: Theory*, Springer Series in Statistics, Probability and its Applications, Springer, New York, etc., 1998.

[147] A. Rahmim, J. Tang, and H. Zaidi, *Four-dimensional (4D) image reconstruction strategies in dynamic pet: Beyond conventional independent frame reconstruction*, Med. Phys. **36** (2009), 3654–3670. 26, 170

[148] S. Remmele, M. Seeland, and J. Hesser, *Fluorescence microscopy deconvolution based on bregman iteration and richardson-lucy algorithm with tv regularization*, Proceedings of the Workshop: Bildverarbeitung für die Medizin 2008, Informatik aktuell, Springer, Berlin, 2008, pp. 72–76.

[149] E. Resmerita, *Regularization of ill-posed problems in banach spaces: convergence rates*, Inverse Problems **21** (2005), 1303–1314. 63

[150] E. Resmerita and R.S. Anderssen, *Joint additive Kullback-Leibler residual minimization and regularization for linear inverse problems*, Math. Meth. Appl. Sci. **30** (2007), 1527–1544. 114, 121, 122, 129

[151] E. Resmerita, H.W. Engl, and A.N. Iusem, *The expectation-maximization algorithm for ill-posed integral equations: a convergence analysis*, Inverse Problems **23** (2007), 2575–2588. 112, 113

[152] E. Resmerita and O. Scherzer, *Error estimates for non-quadratic regularization and the relation to enhancing*, Inverse Problems **22** (2006), 801–814. 63

[153] W.H. Richardson, *Bayesian-Based Iterative Method of Image Restoration*, J. Opt. Soc. Am. **62** (1972), no. 1, 55–59. 105, 111

[154] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton NY, 1970. 51, 57, 59, 84

[155] R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, A Series of Comprehensive Studies in Mathematics, vol. 317, Springer, 1997. 51, 57, 59

[156] R.T. Rockafellar, *Augmented lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research **82** (1956), 421–439. 89

[157] L. Rudin, P. L. Lions, and S. Osher, *Multiplicative denoising and debluring - theory and algorithms*, Geometric Level Sets in Imaging, Vision, and Graphics (S. Osher and N. Paragios, eds.), Springer, New York, 2003, pp. 103–119. 36, 37

[158] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), 259–268. 43, 66, 104, 105, 113, 114

[159] L. Ruthotto, *Mass-preserving registration of medical images*, Master's thesis, Institute for Computational and Applied Mathematics, University of Münster, 2010. 172

[160] A. Sawatzky, C. Brune, J. Mueller, and M. Burger, *Total variation processing of images with Poisson statistics*, Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns (X. Jiang and N. Petkov, eds.), vol. 5702, Springer, Juli 2009, Doi: 10.1007/978-3-642-03767-2_65, pp. 533–540.

[161] A. Sawatzky, C. Brune, F. Wübbeling, T. Kösters, K. Schäfers, and M. Burger, *Accurate EM-TV algorithm in PET with low SNR*, 2008 IEEE Nuclear Science Symposium Conference Record, November 2008, pp. 5133–5137.

[162] K.P. Schäfers, T.J. Spinks, P.G. Camici, P.M. Bloomfield, C.G. Rhodes, M.P. Law, C.S.R. Baker, and O. Rimoldi, O., *Absolute quantification of myocardial blood flow with $H_2{}^{15}O$ and 3-Dimensional PET: An experimental validation*, J. Nucl. Med. **43** (2002), 1031–1040. 157

[163] O. Scherzer and C. Groetsch, *Inverse scale space theory for inverse problems*, Scale-space and morphology in computer vision. Proc. 3rd Int. Conf. Scale-space (M. Kerckhove, ed.), Springer, Berlin, 2001, pp. 317–325. 70

[164] S. Setzer, *Split bregman algorithm, douglas-rachford splitting and frame shrinkage*, Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, X.-C. Tai et al. (eds.), Springer, vol. 5567, 2009, pp. 464–476.

[165] S. Setzer, *Splitting methods in image processing*, Ph.D. thesis, University of Mannheim, 2009. 88, 89, 97, 119

[166] S. Setzer, *Operator splittings, bregman methods and frame shrinkage in image processing*, International Journal of Computer Vision, accepted (2010). 97

[167] S. Setzer, G. Steidl, and T. Teuber, *Deblurring poissonian images by split bregman techniques*, Journal of Visual Communication and Image Representation **21** (2010), 193–199.

[168] L.A. Shepp and Y. Vardi, *Maximum Likelihood Reconstruction for Emission Tomography*, IEEE Trans. Med. Imaging **1** (1982), no. 2, 113–122. 105, 109, 111, 112

[169] R. E. Showalter, *Monotone Operators in Banach Spaces and Nonlinear Partial Differential Equations*, Mathematical Surveys and Monographs, vol. 49, American Mathematical Society, 1997. 220

[170] I. Singer, *Duality for Nonconvex Approximation and Optimization*, Springer, 2006. 54

[171] G. Steidl and T. Teuber, *Anisotropic smoothing using double orientations*, Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision (X.-C. Tai, K. Morken, M. Lysaker, and K.-A. Lie, eds.), LNCS 5567, Springer Verlag - Berlin/Heidelberg, 2009, pp. 477–489. 144

[172] D.M. Strong, J.-F. Aujol, and T.F. Chan, *Scale recognition, regularization parameter selection, and meyer's G-norm in total variation regularization*, Multiscale Model. Simul. **5** (2006), 273–303. 118

[173] X.-C. Tai and C. Wu, *Augmented lagrangian method, dual methods and split bregman iteration for rof model*, The 2nd international conference, SSVM 2009, LNCS 5567, Springer, 2009, pp. 502–513.

[174] A. B. Tayler, *Mathematical Models in Applied Mechanics*, Oxford Applied Mathematics and Computing Science Series, 1986. 205

[175] R. Temam, *Navier-stokes equations: Theory and numerical analysis*, 3 ed., Studies in Mathematics and its Applications, vol. 2, North-Holland Publishing Company - Amsterdam/Oxford, 1984. 83

[176] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim. **29** (1991), 119–138. 93, 94, 106, 119, 132

[177] C. Van Loan, *Computational Frameworks for the Fast Fourier Transform*, SIAM, Frontiers in Applied Mathematics, 1992.

[178] Y. Vardi, L.A. Shepp, and L. Kaufman, *A Statistical Model for Positron Emission Tomography*, J. Am. Stat. Assoc. **80** (1985), no. 389, 8–20. 20, 104, 109, 112, 154

[179] S.S. Vedam, V.R. Kini, P.J. Keall, V. Ramakrishnan, H. Mostafavi, and R. Mohan, *Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker*, Med Phys **30** (2003), no. 4, 505–513. 25, 169

[180] R. Verfürth, *A combined conjugate gradient-multigrid algorithm for the numerical solution of the stokes problem*, IMA Journal of Numerical Analysis **4** (1984), 441–455. 86

[181] J. Verhaeghe, P. Gravel, R. Mio, R. Fukasawa, P. Rosa-Neto, J.-P. Soucy, C.J. Thompson, and A. J. Reader, *Motion-compensated fully 4D PET reconstruction using PET data supersets*, IEEE Nuclear Science Symposium and Medical Imaging Conference, 2009. 26, 170

[182] L.A. Vese and S.J. Osher, *Modeling textures with total variation minimization and oscillating patterns in image processing*, J. Sci. Comput. **19** (2003), 553–572. 148

[183] C. Villani, *Topics in Optimal Transportation*, vol. 58, American Mathematical Society, 2003. 75

[184] ———, *Optimal Transport - Old and New*, A Series of Comprehensive Studies in Mathematics, vol. 338, Springer, Berlin, Heidelberg, 2009.

[185] C.R. Vogel, *Computational Methods for Inverse Problems*, Frontiers in Applied Mathematics, SIAM, Philadelphia, 2002. 118

[186] J. Weickert, *Anisotropic Diffusion in Image Processing*, ECMI Series, Teubner; Stuttgart, 1997. 189, 191

[187] J. Weickert and C. Schnörr, *A theoretical framework for convex regularizers in PDE-based computation of image motion*, International Journal of Computer Vision **45** (2001), no. 3, 245–264. 191

[188] ———, *Variational optic flow computation with a spatio-temporal smoothness constraint*, Journal of Mathematical Imaging and Vision **14** (2001), no. 3, 245–255. 190

[189] M.N. Wernick and J.N. Aarsvold, *Emission Tomography: The Fundamentals of PET and SPECT*, Elsevier Academic Press, 2004. 20, 109, 154

[190] P. Wesseling, *Principles of computational fluid dynamics*, Springer Series in Computational Mathematics, vol. 29, Springer, Berlin, 2001. 83

[191] K.I. Willig, B. Harke, R. Medda, and S.W. Hell, *STED microscopy with continuous wave beams*, Nature Meth. **4** (2007), 915–918. 18, 153, 157

[192] A.P. Witkin, *Scale-space filtering*, Proc. Int. Joint Conf. on Artificial Intelligence, Karlsruhe, 1983, pp. 1019–1023. 70

[193] C. Wu, J. Zhang, and C.-X. Tai, *Augmented lagrangian method for total variation restoration with non-quadratic fidelity*, Tech. report, UCLA CAM Report 09-82, 2009. 87, 89

[194] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for l1-minimization with applications to compressed sensing*, SIAM Journal on Imaging Sciences **1** (2008), no. 1, 143–168. 69

[195] J. Yuan, C. Schnörr, and G. Steidl, *Convex hodge decomposition and regularization of image flows*, J. Math. Imag. Vision **33** (2009), no. 2, 169–177. 197

[196] J. Yuan, G. Steidl, and C. Schnörr, *Simultaneous optical flow estimation and decomposition*, SIAM J. Scientific Computing **29** (2007), no. 6, 2283–2304. 197

[197] E. Zeidler, *Nonlinear Functional Analysis and Its Applications III: Variational Methods and Optimization*, Springer, New York, 1985. 48

[198] X. Zhang, M. Burger, X. Bresson, and S. Osher, *Bregmanized nonlocal regularization for deconvolution and sparse reconstruction*, Tech. report, UCLA CAM Report 09-03, 2009. 69, 87, 89, 98, 99, 100

[199] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on bregman iteration*, Tech. report, UCLA CAM Report 09-99, 2009. 87, 89

[200] Walter Zulehner, *Analysis of iterative methods for saddle point problems: A unified approach*, AMS, Mathematics of Computation **71** (2001), no. 238, 479–505. 87